ESA CONTRACT REPORT

Contract Report to the European Space Agency

Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction

WP-2000 report: Observation quality monitoring and pre-processing

M. Fielding and M. Janisková

ESA ESTEC contract
4000116891/16/NL/LvH

ECMWF

Series: ECMWF ESA Project Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

Contract Report to the European Space Agency

Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction

WP-2000 report: Observation quality monitoring and pre-processing

Authors: M. Fielding and M. Janisková

May 2017

ABSTRACT

This report details the development of a monitoring system and observations pre-processing system in preparation for the direct assimilation of Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE) cloud radar and lidar within the Four-Dimensional Variational Assimilation (4D-Var) system used at European Centre for Medium Range Weather Forecasts (ECMWF). The work is divided into three sections. Firstly, the observation operators and their gradient are refined and tested for inclusion in the Integrated Forecasting System (IFS) of ECMWF. Secondly, initial screening thresholds, bias correction and observation error specified as part of 'scientific pre-processing developments'. Finally, the data monitoring system is defined, using a combination of in-house tools and off-line experiments to select an optimal set of indicators for the detection of instrument drift and errors.

Contents

1   Introduction

The Earth, Clouds, Aerosols and Radiation Explorer (EarthCARE; Illingworth et al., 2015) mission is a unique opportunity to operationally assimilate satellite-based cloud radar and lidar observations into a numerical weather prediction (NWP) model. Cloud related radiances have been used within the ECMWF 4D-Var assimilation system since March 2009 (Bauer et al., 2010; Geer et al., 2010), but profiling observations of clouds, such as those given by EarthCARE, represent a new frontier. While these new observations are an un-tapped resource, potentially providing vertically resolved information to remote, otherwise unobserved parts of the atmosphere, significant developments are required to enable their useful ingestion to the assimilation system.

Fundamentally, the assimilation of each observation within 4D-Var relies on three pieces of information: the observed variable itself, the model equivalent of the observation, and the expected error between the two. The model equivalent is computed by the observation operator, which takes the model state of the atmosphere as input. The amount of information that each observation provides is calculated by the difference between the observed variable and the model equivalent, known as the first guess departure, which is then inversely weighted by the observation error. While appearing simple, the derivation of these steps is far from trivial for cloud radar and lidar observations.

Fortunately the foundations for defining the observation operator and specifying the observation error were made in two previous projects (QuARL, Janisková et al., 2010; and STSE, Janisková et al., 2014), however they need to be finalized and updated for in-line assimilation, where operational constraints require a balance of efficiency and complexity. There are also significant technical developments needed to process the raw observations provided by the European Space Agency (ESA) and The Japan Aerospace Exploration Agency (JAXA). Once the observation handling and pre-processing system is in place, a monitoring system can be implemented. This will provide useful feedback to ESA and JAXA to help detect potential problems with the satellite and instruments. It also provides feedback to the assimilation and forecasting system as a whole, highlighting any model issues that need attention.

This document outlines the observation operator, scientific observation pre-processing and the characterization of observation error required for assimilation of EarthCARE cloud radar and lidar observations. The document also details the necessary developments for including the observations in an automatic monitoring system. The technical developments will be included in WP-3000 (Assimilation system development for cloud radar and lidar observations, Janisková et al., 2017). In Section 2, the definition of the observation operator, which is required to convert model control variables into so-called observation space is given. Section 3 details the specification of appropriate errors that are representative for both the observation and model estimate. Finally in Section 4 the development of monitoring system is discussed. A summary concludes the report in Section 5.

2   Observation operator development

The aim of this section is to bring together previous developments with the latest research to define an efficient, state-of-the-art observation operator ready for inclusion within ECMWF's Integrated Forecast System (IFS). As outlined in the introduction, significant developments were made in the QuARL and STSE projects and here we concentrate on refining both the radar and lidar operators to suit operational requirements and ensure consistency with IFS parameterizations and other observation operators. In this document, the observation operators are tailored to the specification of the CloudSat radar and CALIPSO lidar. The adaptions required for the Cloud Profiling Radar (CPR) and ATmospheric LIDar (ATLID) on-board EarthCARE will be made in WP-4000 (EarthCARE data handling and testing, Fielding et al., 2017).

2.1   Radar observation operator

The radar forward operator developed in this project is ZmVar (reflectivity model for variational assimilation). It has been used previously to simulate observations from ground based and satellite observations, including the precipitation radar on board the Tropical Rainfall Measuring Mission (TRMM) satellite (Benedetti et al., 2006) and the cloud radar on board CloudSat (Janisková et al., 2012). As ZmVar is well documented from previous studies we give only a brief overview of its structure and refer the reader to the aforementioned reports for further details.

Figure 2.1 outlines the steps to convert a model atmospheric profile to simulated observations of radar reflectivity. To save the computational cost of exact calculations on every profile, some calculations are made off-line. Specifically, the bulk microwave scattering properties (backscatter, extinction, single scattering albedo and asymmetry factor) are pre-computed at discrete intervals of water content and temperature by integrating individual particle single scattering properties over a given particle size distribution (PSD). The bulk scattering properties are computed for the six different prognostic hydrometeor types used within the IFS: stratiform and convective rain, stratiform and convective snow and liquid cloud and ice cloud. These are then stored in a lookup table (Step 2).

In its original form, for each model atmospheric profile and level where the model-equivalent radar reflectivity is required, a bilinear interpolation is performed using the model temperature, T, and in-cloud water content, w, to obtain the unattenuated radar reflectivity, Z^j, and extinction, α^j, for each hydrometeor type, j, from the pre-computed lookup table (Step 3). However, as an alternative method with additional computational cost savings, the relationship between Z, T and w can be parameterized using a two-variable two-degree polynomial with six fitted coefficients (a^i):

$$Z^j = a_0 + a_1 w + a_2 T + a_3 w^2 + a_4 T^2 + a_5 wT. \tag{2.1}$$

A similar expression is used to approximate α^j. Further details of the fitting method and computational savings are given in section 2.4.2.

The total unattenuated radar reflectivity is simply a sum of each hydrometeor type, i.e.:

$$Z = \sum_{j=1}^{N} Z^j \tag{2.2}$$

where N is the number of hydrometeor types.

Attenuation of the signal along the radar beam must also be taken into account. Clear sky attenuation due to

gases (Step 4) is calculated using the models of Liebe (1985) and Liebe et al. (1992), and cloudy sky attenuation (Step 6) is given by the sum of clear sky attenuation and the hydrometeor extinction, i.e.:

$$\alpha = \alpha^{gas} + \sum_{j=1}^{N} \alpha^j \tag{2.3}$$

ZmVar has three methods for implementing the attenuation (see Sec. 2.3). In its most basic, 'single-column' form, ZmVar computes the attenuated reflectivity, Z_l^a as:

$$Z_l^a = Z_l e^{-2\tau_l} \frac{1 - e^{-2\Delta h_l \alpha_l}}{2\Delta h_l \alpha_l} \tag{2.4}$$

where Δh_l is the depth of the lth layer and τ is the optical depth of all the previous layers between the lth layer and the instrument.

In addition to the single scattering properties given by the look-up table, the original version of ZmVar also includes a model for multiple scattering. Full details on the implementation are given in Di Michele et al. (2014a), so only a short summary is provided here. In the single-scattering assumption we assume that radiation scattered in any direction apart from the exact backscatter direction can be neglected. Multiple scattering therefore becomes important when the optical thickness of the medium is sufficient to cause a significant additional backscatter arising from previously scattered radiation that has remained or re-entered the field of view of the instrument. The validity of the single scattering assumption for cloud radar such as CloudSat or the CPR on-board EarthCARE only breaks down for heavy precipitation, such as within convective cores (Battaglia et al., 2011).
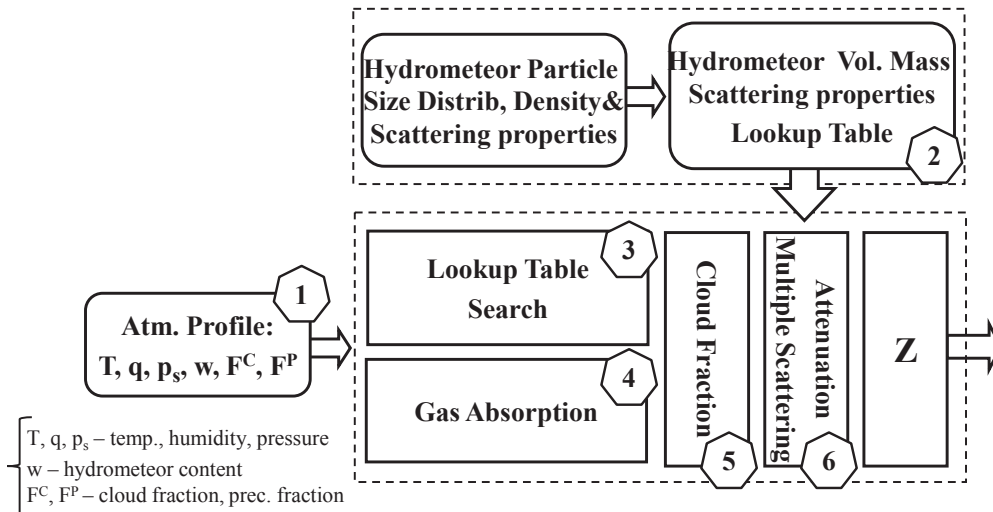
Figure 2.1: Schematic diagram of ZmVar operator as shown in Di Michele et al. (2014a). The numbers indicate the order of the operator computations.

To account for occasions where multiple scattering is non-negligible, ZmVar includes the option of running the time dependent two-stream (TDTS; Hogan and Battaglia, 2008) scheme. The method involves tracking the propagation of radar pulses in time as well space, and is therefore relatively costly to run compared to running ZmVar with just the single-scattering approximation. A more detailed comparison of the costs is given in Sec. 2.4. However, at present, the multiple scattering option is only available for evaluation and monitoring; for data assimilation, situations where multiple scattering is likely are excluded.

Changes to the microphysical and scattering assumptions in the forward operator

Any changes to the microphysical and scattering assumptions made in ZmVar will affect the bulk unattenuated backscatter and extinction properties stored within the pre-computed look-up table. As discussed above, for each model hydrometeor type there are three main selections to be made: the particle size distribution (PSD), the particle shape and the scattering model that provides the single-particle scattering properties. Integrating the single scattering properties over the chosen PSD results in the required bulk hydrometeor scattering properties.

Much effort was spent in the previous projects optimizing the microphysical assumptions to first guess departures of CloudSat and CALIPSO observations. However, we must be careful not to 'over-tune' our assumptions to a particular instrument or model cycle, especially if the assumptions are unphysical. Consistency in the assumptions made by the operators and IFS model is also important; a lack of consistency increases the possibility of the operators 'pulling in different directions' during the 4D-Var minimization, which could lead to sub-optimal analysis increments. Therefore, the emphasis of the changes is consistency and physical realism.
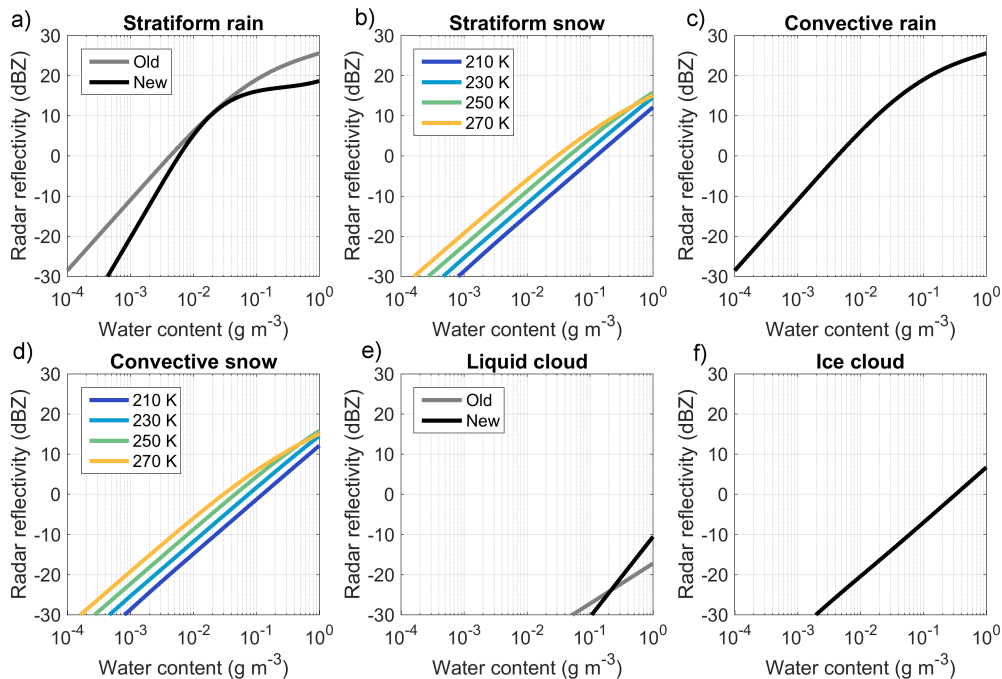
Figure 2.2: Relationship between water content and simulated unattenuated radar reflectivity as prescribed in the ZmVar lookup table.

The updated microphysical assumptions for the radar forward operator are given in Table 2.1. The corresponding relationships between hydrometeor water content and unattenuated radar reflectivity are visualised in Fig. 2.2. Note, as in the previous project, stratiform snow and convective snow are treated using the same assumptions. Also, the ice cloud bulk scattering properties are independent of temperature to force the PSD to be shifted towards smaller particles for all temperatures and water contents.

There are two main updates to the radar look-up table:

- We adopt the Abel and Boutle (2012) PSD for large scale rain.
- The lognormal PSD for cloud liquid water uses a fixed droplet number concentration rather than a fixed droplet radius.

Hydrometeor	Particle shape	PSD and parameters	Density $[g/cm^3]$ $\rho(D) = aD^b$	
Cloud Liquid	Sphere	$n(r) = \frac{N_t}{\sqrt{2\pi}(\ln\sigma_g)r}e^{-\frac{\ln^2(r/r_g)}{2(\ln\sigma_g)^2}}\begin{cases} N_t = 100cm^{-3} \\ \sigma_g = 710^{-4} \end{cases}$	$a = 1$	$b = 0$
Cloud Ice	6-bullet rosette	*Field et al. (2007), no Temp dependence*	$a = 9.40\ 10^{-3}$	$b = -0.87$
Rain LS	Sphere	$N(D) = N_0\exp(-\lambda D), N_0 = x_1\lambda^{x_2}\begin{cases} x_1 = 0.22 \\ x_2 = 2.2 \end{cases}$	$a = 1$	$b = 0$
Snow LS	Aggregate of columns	*Field et al. (2007)*	$a = 2.60\ 10^{-3}$	$b = -1.42$
Rain Conv	Sphere	$N(D) = \frac{0.03N_L D_0^4\Lambda^{\mu+4}}{\Gamma(\mu+4)}D^\mu e^{-\Lambda D}\begin{cases} N_L = 0.08cm^{-4} \\ \Lambda = \frac{3.6+\mu}{D_0} \end{cases}$	$a = 1$	$b = 0$
Snow Conv	Aggregate of columns	*Field et al. (2007)*	$a = 2.60\ 10^{-3}$	$b = -1.42$

Table 2.1: Revised parametrization of hydrometeor particle properties in ZmVar.

Both of these changes ensure consistency with the IFS cloud scheme; in a recent IFS cycle a new rain evaporation formulation, including fall speed assumptions, was implemented based on Abel and Boutle (2012). Abel and Boutle designed their scheme specifically for single-moment microphysics schemes, as used in the IFS. By parameterizing the number concentration as a function of liquid water content, the new scheme better represents both the smaller rain drops typically found in drizzling boundary layer cloud and the transition to larger drops found in heavier precipitation.

Figure 2.2a shows the effect of the new PSD on the radar reflectivity for stratiform rain. The new scheme's shift towards smaller drops at lower water content results in a significant reduction in radar reflectivity for water content less than $0.01\,\mathrm{g\,m^{-3}}$. Conversely, the shift towards larger drops for water content greater than $0.01\,\mathrm{g\,m^{-3}}$ also results in a decrease in radar reflectivity due to more drops undergoing Mie- rather than Rayleigh-scattering.

The second update of using a fixed droplet number concentration for cloud droplets also brings consistency with the IFS cloud scheme, where cloud droplet number concentration is constant within a grid-box for calculating the autoconversion/accretion of cloud water to rain. For simplicity we choose a globally fixed value of $100\,\mathrm{cm^{-3}}$, whereas in the IFS a value of $50\,\mathrm{cm^{-3}}$ is used over ocean and $300\,\mathrm{cm^{-3}}$ is used over land. Assuming a fixed droplet number concentration is also both consistent with observations (e.g., Miles et al., 2000) and simple theoretical arguments; in an adiabatic parcel model of a cloud, cloud droplets are activated at cloud base and rise and grow in up-drafts, effectively fixing their number concentration. The update only has a small effect on the simulated radar reflectivity; Fig 2.2e shows that the gradient of radar reflectivity with water content is greater for the updated scheme due to the increase in cloud droplet size for greater water content.

The combined effect of these changes can be seen in the frequency distributions of simulated radar reflectivity with temperature as shown in Figure 2.3, where the updated scheme (Fig. 2.3c) shows a greater proportion of occurrences of radar reflectivity less than -20 dBZ. Qualitatively, this is a closer match to the observed (Fig. 2.3a) distribution. Nevertheless, both schemes (old and new) show a remarkable resemblance to the observations, accrediting both IFS model and ZmVar developments.
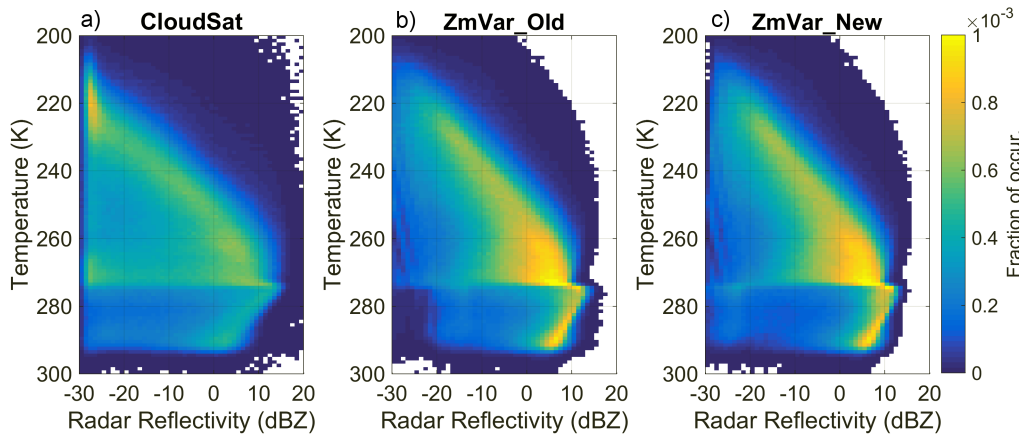
Figure 2.3: Frequency distribution of observed and simulated radar reflectivity with temperature. Panel (a) shows Cloud-Sat observations for August 2007 after averaging at model resolution. Panels (b) and (c) show the simulated reflectivity using the original and updated lookup tables.

2.2 Lidar observation operator

The basic principle of lidar is much the same as radar, except that the typical wavelength of lidar ($\sim 1\,\mu m$) is much shorter than radar ($\sim 1\,cm$). This leads to two additional processes that ZmVar must account for to provide realistic simulated lidar observations. The first is scattering from gases, known as Rayleigh scattering or molecular backscatter. Molecular backscatter is straightforward to model using the following expression (Collis and Russell, 1976):

$$\beta_\lambda^{mol} = 5.45 \cdot 10^{-32} \times \frac{p}{KT} \times \left(\frac{\lambda}{0.55} \right)^{-4.09} \tag{2.5}$$

where p is the atmospheric pressure, K is the Boltzmann constant ($1.38 \cdot 10^{-23} J K^{-1}$) and λ is the lidar wavelength.

The second is that the lidar photons are much more likely to undergo multiple scattering within clouds than radar due to greater extinction and strong forward peak of the phase function. ZmVar for lidar has two options for modeling multiple scattering; either the Photon Variance-Covariance (PVC; Hogan, 2008) method or the 'Platt approximation' (Platt, 1973). The PVC is a precise and efficient method to account for the majority of multiple scattering, and it uses the exact specification of the lidar and geometry of the satellite position. Alternatively, the simpler Platt approximation uses a multiplier, η, which is applied to the optical depth of each cloudy layer. The value of η can vary between 1 (the single-scattering limit) and 1/2 (the wide field-of-view limit). In the STSE project, η was tuned using the PVC method as a function of temperature. The relative performance of the two methods is briefly discussed in Sec. 2.4. Apart from these additions, ZmVar for lidar follows the same procedure as described in Section 2.1. Full details can be found in Di Michele et al. (2014b).

Changes to the microphysical assumptions in the forward operator

As discussed for the radar observation operator, the changes to the microphysical assumptions are a balance of: ensuring consistency between the IFS model and other operators; minimizing FG departures of existing observations; and maintaining their 'physical realism' by basing the operator assumptions on known hydrometeor properties and processes.
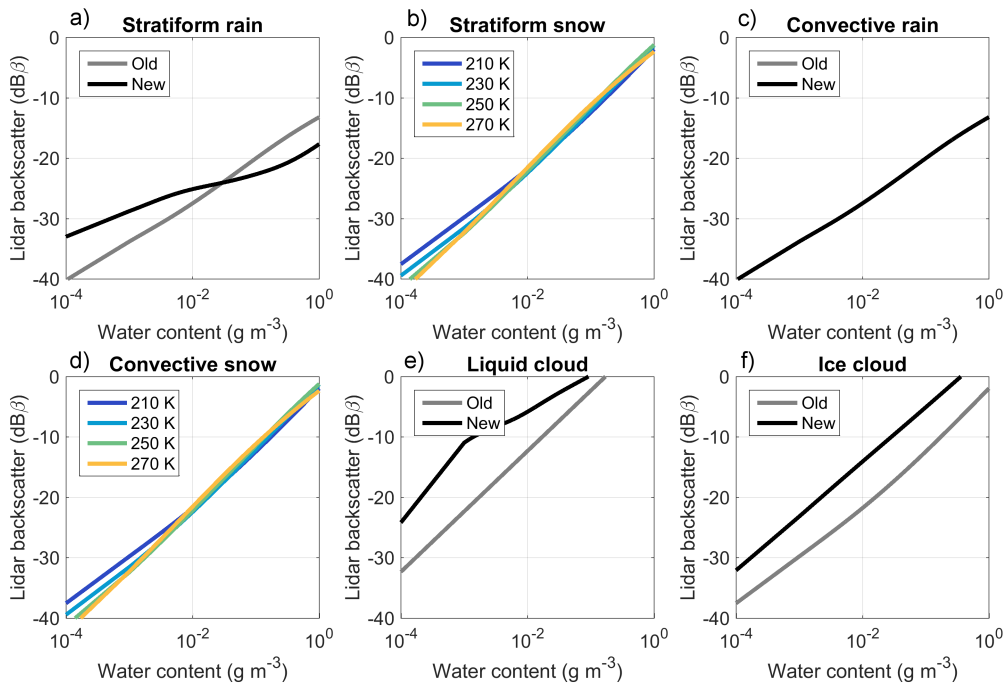
Figure 2.4: Relationship between water content and simulated unattenuated lidar backscatter as prescribed in the ZmVar lookup table. Note the old (grey) and new (black) tables are identical for convective rain.

The hydrometeor microphysical properties assumed for lidar are now identical to the assumptions for radar (Table 2.1). This equates to three updates from the previous project:

- We adopt the Abel and Boutle (2012) PSD for large scale rain.
- The lognormal PSD for cloud liquid water uses a fixed droplet number concentration rather than a fixed droplet radius.
- The ice cloud particle type is changed from 'aggregate of columns' to '6-bullet rosette'.

The impact of these changes is shown in Fig. 2.4. The new PSD for stratiform rain (Fig. 2.4a) reduces the sensitivity of backscatter to changes in rain water content. For smaller water content, the reduced drop sizes given by the new scheme increase the lidar backscatter. For larger water content, the larger drop sizes prescribed by the scheme relative to the old normalised gamma PSD reduce the backscatter. For liquid cloud (Fig. 2.4e), fixing the droplet number concentration at $100 \, \text{cm}^{-3}$ gives smaller droplet sizes for water content less than $0.1 \, \text{g m}^{-3}$. Finally, using the '6-bullet rosette' single scattering properties from the Yang et al. (2000) database increases the lidar backscatter relative to the 'rough aggregate' that was used previously.

A further evaluation of the microphysical assumptions for both radar and lidar is given in Sec. 3.2 as part of the updates to the bias correction scheme.

2.3 Cloud overlap assumptions

Despite recent increases in model resolution, with the current operational high resolution IFS forecast implementing a grid spacing of about 9 km, the variability of cloud and precipitation within a grid column should be taken into account. In addition to the grid box average water content, the IFS outputs a prognostic sub-grid cloud fraction, which is used to define the in-cloud water content, w'. While it is straightforward to define w'

for each model level, the way clouds and precipitation overlap between levels is not obvious, yet it can have a significant impact on the model-equivalent attenuated signal. The choice of overlap assumption, and method to implement it, is particularly important for lidar, or radar where there is strong attenuation such as in deep convection.

Cloud overlap has long been recognised as important for accurate radiative transfer in weather and climate models (Liang and Wang, 1997). In the literature, there are four main overlap types: minimum, maximum, random (Geleyn and Hollingsworth, 1979; Morcrette and Fouquart, 1986) and exponential (Hogan and Illingworth, 2000). All of these types can be used to define the joint areal fraction, C_{ij}, between two vertically adjacent cloud layers, C_1 and C_2. The first three overlap types are independent of cloud depth:

$$\text{– minimum overlap (MIN):} \quad C_{ij}^{\text{MIN}} = \min\{C_1, C_2\} \tag{2.6a}$$

$$\text{– maximum overlap (MAX):} \quad C_{ij}^{\text{MAX}} = \max\{C_1, C_2\} \tag{2.6b}$$

$$\text{– random overlap (RAN):} \quad C_{ij}^{\text{RAN}} = C_1 + C_2 - C_1 C_2, \tag{2.6c}$$

whereas exponential overlap assumes that thin cloud layers are near maximally overlapped, relaxing to random as the distance, D, between layers increases:

$$\text{– exponential overlap (EXP):} \quad C_{ij}^{\text{EXP}} = \alpha C_{ij}^{\text{MAX}} + (\alpha - 1) C_{ij}^{\text{RAN}} \tag{2.7}$$

with $\alpha = exp(-D/L)$, where L is the vertical decorrelation length. Hogan and Illingworth (2000) showed using cloud radar observations that for continuous cloud, L was around 2 km. For non-continuous cloud, they found that the correlation decreased sharply as the vertical separation increased. For large separations between cloud layers, they found α tends to zero, equivalent to assuming random overlap. Operationally, many NWP and climate models assume exponential-random (EXP-RAN) overlap. It is also computationally convenient to assume that information of relative cloud position within a grid box is lost between clouds layers that are not contiguous.

The previous version of ZmVar has two approaches for handling sub-grid cloud and precipitation overlap (described in full in Di Michele et al., 2014a) . The first, single-column approach, assumes that the hydrometeor optical depth of each layer is proportional to the in-cloud extinction multiplied by the cloud fraction. As the extinction coefficient scales quasi-linearly with hydrometeor content, the single-column approach effectively assumes each layer is overcast. The main advantage of this approach is that it is computationally efficient, but it is inflexible in that it cannot be customized with a particular overlap assumption and therefore is unlikely to provide the best representation of clouds and precipitation compared to reality.

In contrast, the second, multi-column approach, is based on the Sub-grid Cloud Overlap Profile Sampler (SCOPS; Webb et al., 2001) and is completely flexible to assume any overlap assumption. It works by creating a set of sub-columns, where each level is assigned to be either totally cloudy or totally clear, such that the sum of sub-columns at each level is equivalent to the grid cloud fraction, and their distribution is consistent with an assumed overlap scheme. The effectiveness of this approach depends on the number of sub-columns, N_{col}, as it is clear that the smallest discrete cloud fraction that can be resolved depends on $1/N_{col}$. While using the multi-column approach with 50+ sub-columns may lead to more realistic computations than the single-column approach, clearly the additional computational cost is significant and, if not prohibitively expensive, likely better spent elsewhere.

A new double-column approach

Here we develop a new double-column approach that borrows ideas from the NWP radiation schemes (e.g., the Edward Slingo radiation scheme (ESRAD); Edwards and Slingo, 1996). The method computes the transmission

of an active sensor signal as it passes through the atmosphere by, at each model level, partitioning the signal into a uniform cloudy layer and a uniform clear layer. As in ESRAD, the proportion of signal that travels in and out of either the cloudy or clear layer is governed by an assumed overlap. Unlike the single-column approach, it can account for different overlap assumptions, while being much more computationally efficient than the multi-column approach. We now describe the method in detail, before comparing its performance against the other approaches.

The method works by computing the cloudy transmission, T_w, and the clear-sky transmission, T_c, iteratively through the atmosphere from the top model level to the lowest model level. For each layer, the transmittance is equal to the incoming signal multiplied by the transmissivity of the current layer. The proportion of signal that travels in and out of either the cloudy or clear layer is governed by the assumed overlap between adjacent layers, i.e.:

$$
\begin{pmatrix} T_c^{l+1} \\ T_w^{l+1} \end{pmatrix} = \boldsymbol{O} \begin{pmatrix} \frac{1}{1-C_{l+1}} e^{-2\alpha_c \Delta h_l} \\ \frac{1}{C_{l+1}} e^{-2\alpha_w \Delta h_l} \end{pmatrix} \cdot \begin{pmatrix} T_c^l \\ T_w^l \end{pmatrix},
\tag{2.8}
$$

where the overlap matrix, $\boldsymbol{O}$, is defined as

$$
\boldsymbol{O} = \begin{pmatrix} F_{cc} & F_{cw} \\ F_{wc} & F_{ww} \end{pmatrix} := \begin{pmatrix} 1 - C_{ij} & C_{ij} - C_{l+1} \\ C_{ij} - C_l & C_l + C_{l+1} - C_{ij} \end{pmatrix},
\tag{2.9}
$$

where F_{cc} is the fraction of signal traveling from the clear column to the clear column below, F_{cw} is the fraction traveling from clear to cloudy, F_{wc} is the fraction of signal traveling from cloudy to clear and F_{ww} is the fraction of signal traveling from cloudy to cloudy. By definition, these fractions can be calculated from the two adjacent cloud layers, C_l and C_{l+1}, and C_{ij}. Note that in Eq. 2.8, the attenuation in each layer must be normalised by the respective layer fraction to ensure the transmission in each column is reduced by the correct amount. As in the single- and multiple-column approaches, we also assume that the forward and return paths of the active sensor signal are identical so that the two-way extinction is twice the forward path extinction.

Finally, the grid-box average attenuated reflectivity at any level is simply:

$$
Z_{at}^l = T_c^l Z_c^l + T_w^l Z_w^l.
\tag{2.10}
$$

An analogous expression is used to find the grid-box average attenuated backscatter.

Comparison of overlap methods

To validate the new double-column method we compare its performance against the existing single- and multi-column approaches. The model setup and observation processing is as described in Section 3, so only a brief summary is given here. A set of 12-hour IFS model forecasts using cycle 43r1, initialized with the historic operational analysis and output every 30 minutes, are used to create the input required for ZmVar. CloudSat radar reflectivity and CALIPSO attenuated backscatter observations are matched in space and time to the nearest model grid-box. All observations that fall within a given model grid-box are averaged to create so-called 'superobs'.

In the following plots and discussion the observations and model equivalents span one day (August 1st 2007). Some screening of the observations is applied: only points where both observations and model have cloud fraction greater than 0.1 are considered and observations below 1km are discarded. Model equivalent radar

reflectivity and lidar backscatter are not included where their values are below instrument noise (for simplicity we use fixed values of $-30\,\mathrm{dBZ}$ and $-50\,\mathrm{dB}\beta$ for radar and lidar respectively).
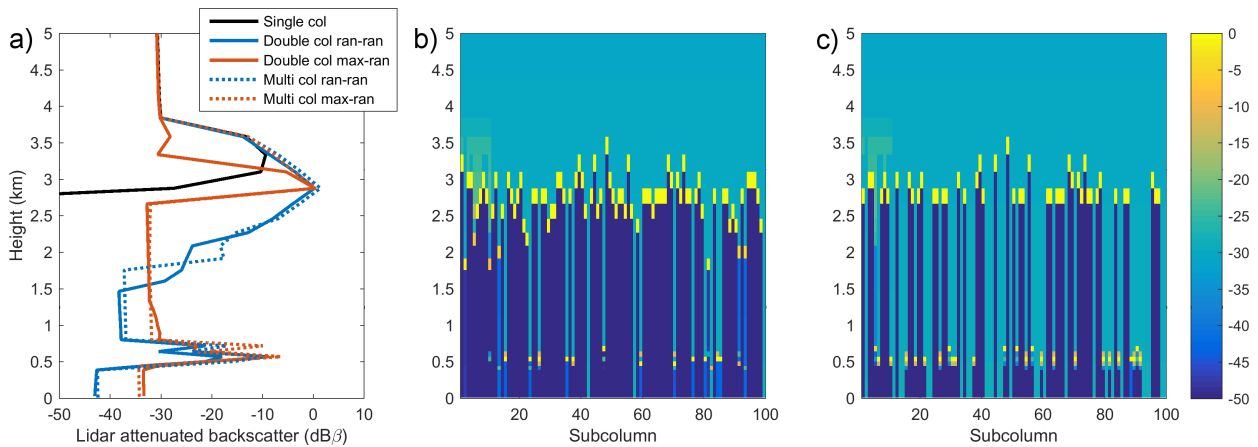

Figure 2.5: (a) Forward modelled lidar attenuated backscatter using the three different solvers available in ZmVar: single column (black line), double-column (solid lines) and multi-column (dashed lines). Panels (b) and (c) show the corresponding sub-columns of lidar backscatter given by the multi-column approach assuming random and max-ran overlap respectively.

Both double- and multi- column methods are set to assume MAX-RAN overlap, that is adjacent cloud layers are maximally overlapped, whereas non-contiguous clouds are assumed to be randomly overlapped. In practice, for the double-column approach, this means setting $\alpha = 1$ (see Eq. 2.7). Note that the double-column implicitly assumes the overlapping of non-contiguous cloud layers is random. We choose 50 independent columns to represent the overlap configuration in the multi-column method; this is chosen as a reasonable balance between computational cost and accuracy. Increasing the number of columns was not found to significantly alter the results.

An example of the attenuated lidar backscatter for the different overlap methods on a single model profile is shown in Fig 2.5. The model profile contains two distinct layers of cloud; the first situated at an altitude of 3 km, the second at 0.5 km. Both layers have cloud fraction of around 0.5. Because the single column method homogonises extinction across the whole model gridbox, the lidar signal is quickly attenuated by the upper cloud layer, such that the lower cloud layer is not detected. Both the double- and multi-column methods allow for some signal to reach the lower cloud layer. Assuming random overlap (Fig. 2.5b) tends to make the upper layer more reflective, whereas MAX-RAN overlap gives a stronger return from the lower cloud layer (Fig. 2.5c).

Figure 2.6 shows the observation minus background (O-B) bias and standard deviation for CloudSat observations and the three different methods. In both the bias and standard deviation we can see that the three methods are almost identical above 5 km. At radar frequencies, attenuation due to hydrometeors is negligible for frozen hydrometeors. Neglecting the multi-column approach's finite sample size, the unattenuated reflectivity, and therefore the O-B statistics, should be identical above the melting layer for all methods.

Below the melting layer, where liquid water can cause significant attenuation, the methods show some differences. Both the double- and multi-column schemes have reduced bias and standard deviation compared to the single column method, suggesting an increase in skill at simulating the observations. In particular the simulated radar reflectivities tend to be relatively larger in the single-column approach. Unlike the double and multi column methods, the single column approach averages the in-cloud extinction to the whole grid box. For cloud fraction less than one, this reduces the overall optical depth within the model column compared to if the layers were maximally overlapped, and hence leads to greater simulated reflectivities.
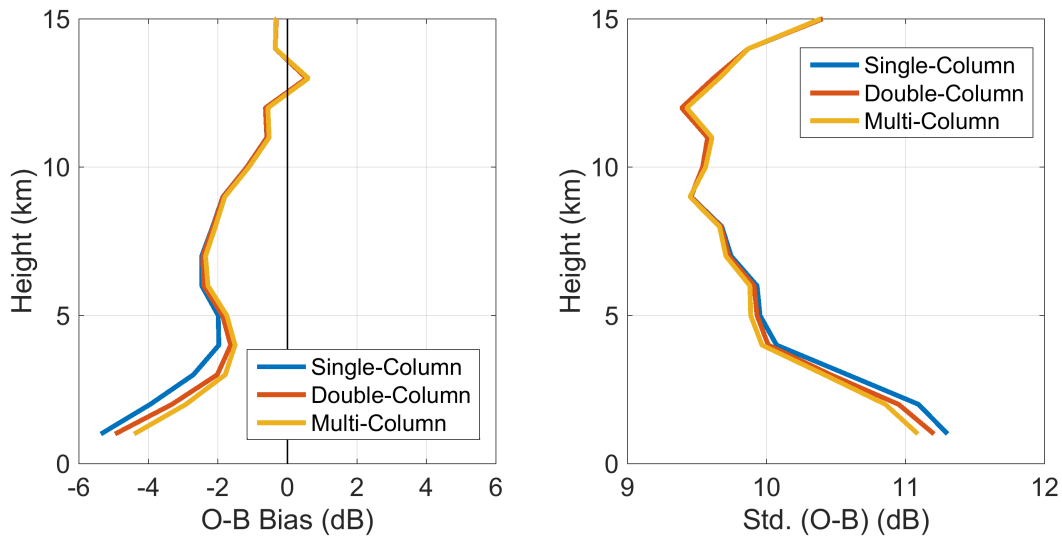
Figure 2.6: Bias (left panel) and standard deviation (right panel) of one days (1st August 2007) CloudSat radar reflectivity versus ZmVar model equivalent using three different methods for handling cloud overlap, stratified by model height. The double-column (red) and multi-column with 50 columns (yellow) both assume MAX-RAN overlap. See Sec. 3 for details of the model setup.
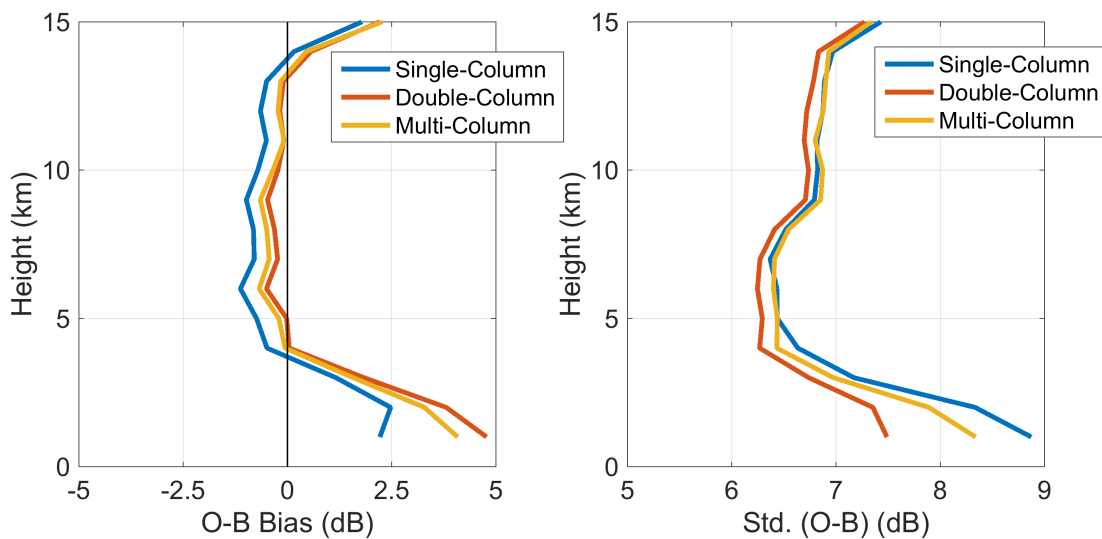

Figure 2.7: Same as Figure 2.6, but using CALIOP attenuated backscatter.

The differences between the methods are greater for CALIOP lidar observations (Fig. 2.7) because attenuation is typically much stronger and occurs in both liquid and frozen hydrometeors. Attenuation is greatest in liquid cloud droplets, which can attenuate a lidar signal to below the instrument sensitivity within a few hundred metres. Contrary to the CloudSat observations, the single-column approach tends to underestimate the gridbox average attenuated lidar backscatter.

The double- and multi-column methods both outperform the single-column approach in both bias and standard deviation. Although the single-column method typically reduces the average optical depth in a particular grid-box column relative to the other methods, it does not allow for a separate clear sky signal to propagate to

lower layers. As a thought experiment to explain this, imagine an optically thick, but broken cloudy layer overlying a completely overcast cloud deck. In the single column method, the lidar signal would be completely attenuated by the upper cloud layer such that the backscatter from the lower cloud layer was undetectable by the instrument. Clearly, in reality some lidar pulses would 'miss' the first broken-cloud layer and reach the lower cloud deck, providing a detectable signal.

Interestingly, the double-column and multi-column approaches show some differences in bias, particularly below 4 km, despite assuming the same overlap. One key difference between the methods is that the double-column approach assumes one cloud fraction for all hydrometeor types within the cloudy column, whereas the multi-column approach can represent any number of hydrometeor types. In practice, this means the hydrometeor properties of a model layer are distributed to match the hydrometeor type with the greatest cloud fraction.

Although the double-column approach assumes a universal hydrometeor fraction for each layer, it is a good approximation to the multi-column approach at a fraction of the computational cost. In fact there is little additional overhead in computational time for the double-column approach versus the single-column approach; the total CPU time per 1000 columns is only 10% more. There is also the potential to expand the method to account for subgrid heterogeneity. Following Shonk and Hogan (2008), a third column could be added to the double column approach to allow two cloudy columns with different microphysical properties.

2.4   Computational cost savings

For operational data assimilation, the observation operators and their tangent linears and adjoints must be kept as efficient as possible. The ECMWF 4D-Var data assimilation system uses an incremental approach in which several minimizations of the cost function are performed at different model resolutions. The highest possible resolution is used for the computation of model trajectory and departures between the observations and model. A lower-resolution version of the model (its adjoint and tangent linear) is used for the iterative and relatively costly computation of analysis increments. Each minimization requires many calls to the tangent linear and adjoint, so computational costs of the observation operators can soon become significant. In this section we detail the computational savings achieved by restructuring the original ZmVar code and implementing a new parameterized look-up table.

2.4.1   Code efficiency

The computational cost of each subroutine within ZmVar have been minimized by optimising loop order, reducing array sizes and removing unnecessary computations. Figure 2.8 gives a breakdown of the different savings by task. The greatest percent gains in efficiency were made in the initialization routines. The most expensive task is computing the radar multiple scattering. As explained in Sec. 2.1, radar multiple scattering is calculated using the TDTS method, which has been developed separately from ZmVar, primarily for use in retrieval algorithms. Although retrievals are typically performed without strict time-constraints, the code was developed for a variational retrieval method that requires several calls to the forward model per retrieved profile, so the method has already been optimised to reduce its cost. As the computational cost is around 0.5 s per column, even if the time taken was halved, the cost would be prohibitive for operational use. The option to use the TDTS method therefore remains in the code for benchmarking and off-line model evaluation experiments.

Whereas the TDTS method is unaffordable for real-time data assimilation, the PVC method for narrow-angle multiple scattering is affordable for lidar applications and the computational cost is only twice that of the single scattering. However, the adaptive Platt method, when calibrated to the PVC method, gives a similar performance with differences in O-B bias and standard deviation less than 1 dB at all model heights evaluated
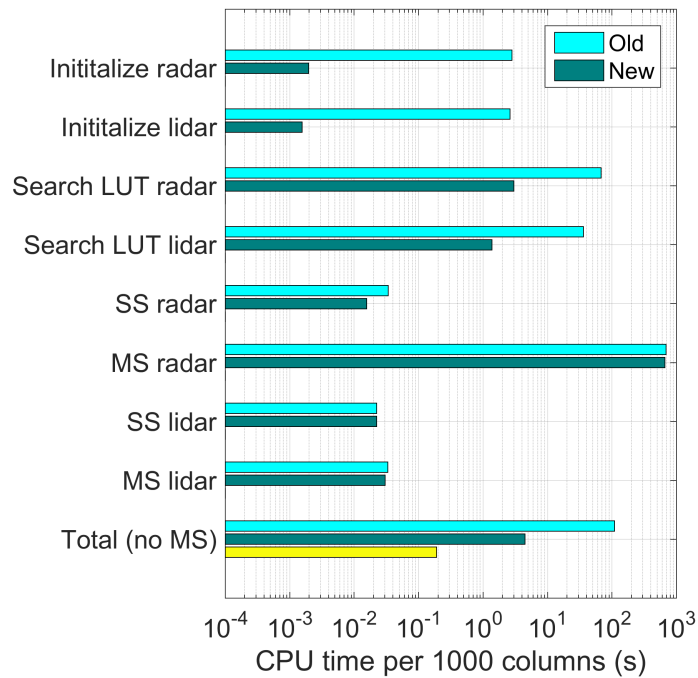
Figure 2.8: Computational cost savings for optimised ZmVar (new) versus the original (old) version, including single scattering (SS) and multiple scattering (MS) options. The yellow bar shows the total computational cost using the parameterized look-up table, see Section 2.4.2 for details. The experiments were conducted on a single Intel(R) Core(TM) i7-4770 CPU at 3.40 GHz.

over a one month period. Given that the coding of the tangent linear and adjoint of the PVC method would require substantial man-hours, we retain the Platt-approximation as the preferred method to account for narrow-angle multiple scattering for the operational data assimilation.

Apart from radar multiple scattering calculations, the most expensive calculations involve searching the lookup tables. Significant savings were made by only searching the look-up tables where hydrometeor mass exceeds a threshold value, and by reducing the sizes of arrays passed between subroutines. In total, the total CPU time per 1000 columns, for both radar and lidar was reduced from around 100 s to less than 10 s. This equates to 60 s CPU time per twelve hours of CloudSat/CALIPSO model equivalents per processor.

2.4.2 Parameterizing the look-up table

One way to reduce the computational cost of searching a pre-computed look-up table is to parameterize the look-up table's output by its input. This is only possible if a suitable 'model' or function can be found. Fortunately, the relationship between water content and radar reflectivity or lidar backscatter can be approximated by simple functions; for most hydrometeors, the relationship between water content and backscatter are quasi-linear if fitted in log-space and so could be fitted with just two parameters. However, to capture some of the non-linear behaviour (in log-space) associated with the transition from Rayleigh scattering to Mie-scattering seen in radar reflectivity for precipitating hydrometeors, a two-variable two-degree polynomial with six fitted parameters (Eq. 2.1) is chosen to model the pre-computed look-up table. To find the coefficients, a non-linear regression is performed, where each entry in the look-up table is considered to be equally-weighted 'observation'. In the current version of the look-up table, there are 406 discrete water content and 70 temperature values.

Figures 2.9 and 2.10 show the agreement between the parameterization and look-up tables for radar reflectivity and lidar backscatter respectively. For radar, the root mean square error (rmse) is well below the expected observation error (around 6 dB, see Sec. 3.3.2) for all hydrometeor types. The worst fit (in terms of rmse) is found for stratiform rain due to the non-linearities in the PSD parameterization and the change from Rayleigh to Mie-scattering for larger drop sizes and water contents. However, the smoother relationship between water content and radar reflectivity given by the parameterization may actually help the 4D-Var minimization to converge, therefore being of benefit to the data assimilation.
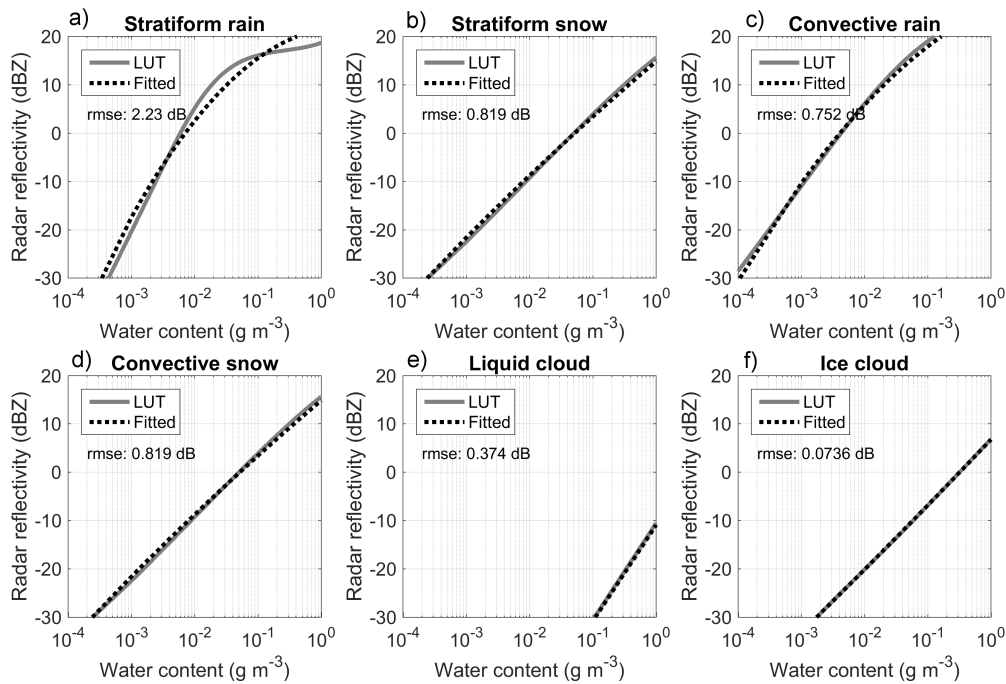

Figure 2.9: *Comparison between look-up table (solid grey line) and parameterized values (black dashed line) as a function of water content for six different hydrometeor types and fixed temperature of 293K for liquid hydrometeors and 253K for solid hydrometeors. The root mean square error (rmse) between look-up table values and the fitting function for all water content and temperatures stored in the look-up table is also shown.*

The same is true when considering the parameterization of lidar backscatter. As the lidar wavelength falls into the Mie-scattering regime for all water contents, the pre-computed look-up table curves tend to be more linear than for the radar and hence have lower rmse. One exception is for liquid cloud, where the smaller droplets found for smaller water contents lead to noisier backscatter values. Again, the parameterization may assist the assimilation in converging and give an overall benefit to the assimilation of lidar observations. Although not shown, the agreement for extinction is similarly reasonable for both radar and lidar.

The computational cost savings of using the parameterized scattering properties is significant. Compared to searching the look-up table, the CPU time for using the parameterization is an order of magnitude less (yellow bar in Fig. 2.8; the original setup costs 4 s per 1000 columns, whereas using the parameterization is less than 0.2 s per 1000 columns. These cost savings allow more time to be allocated to other observation operators or physical parameterizations, while having a minimal impact on accuracy.
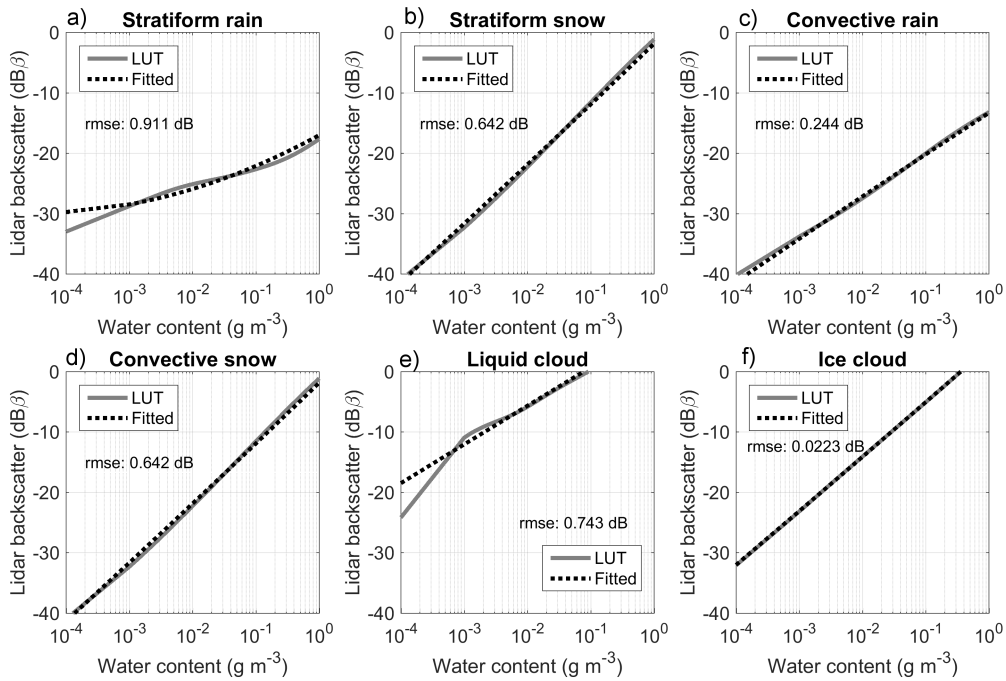
Figure 2.10: Same as Fig. 2.9, but for lidar backscatter.

2.5 Testing of the tangent linear and adjoint codes

In 4D-Var data assimilation system, solving the minimization problem requires the linearization of both the forecast model and the observation operators translating the model control variables (typically temperature, humidity, wind and surface pressure) into some equivalent to assimilated observations. Thus tangent-linear (TL) and adjoint (AD) versions of observation operator for cloud radar reflectivity and lidar backscatter need to be developed.

2.5.1 Tangent-linear version

To build the TL model, the linearization is performed with respect to the local tangent of the model trajectory. If M is the model describing the time evolution of the model state $\mathbf{x}$ at any time t_i as:

$$\mathbf{x}(t_{i+1}) = M[\mathbf{x}(t_i)] \tag{2.11}$$

then the time evolution of a small perturbation $\delta \mathbf{x}$ can be estimated to the first order approximating by the tangent linear model $\mathbf{M}$ (derived from the NL model M):

$$\begin{aligned} \delta \mathbf{x}(t_{i+1}) &= \mathbf{M}[\mathbf{x}(t_i)] \delta \mathbf{x}(t_i) \\ \delta \mathbf{x}(t_{i+1}) &= \frac{\partial M[\mathbf{x}(t_i)]}{\partial \mathbf{x}} \delta \mathbf{x}(t_i) \end{aligned} \tag{2.12}$$

The verification of the correctness of the TL model is first performed through the classical Taylor formula:

$$\lim_{\lambda \to 0} \frac{M(\mathbf{x} + \lambda \, \delta \mathbf{x}) - M(\mathbf{x})}{\mathbf{M}(\lambda \, \delta \mathbf{x})} = 1 \tag{2.13}$$

This examination of asymptotic behaviour, using perturbations the size of which becomes infinitesimally small, is performed to check the numerical correctness of the TL code.

For practical applications (Janisková and Lopez, 2013), an investigation of the accuracy of TL models must be also done for finite-amplitude perturbations (typically perturbations of the size of analysis increments). Only the valid linearized approximation for such perturbations can provide useful results for applications of TL and AD models. Therefore, for the validation of the TL approximation, the accuracy of the linearization of observation operator needs to be studied with respect to pairs of non-linear (NL) results. The difference between two NL integrations (starting from two different initial states, such as a background field, $\mathbf{x}^b$ and an analysis, $\mathbf{x}^a$) run with the NL model (observation operator), M, is compared to evolution of the increments $(\mathbf{x}^a - \mathbf{x}^b)$ obtained by integrating the TL model, $\mathbf{M}$. The results from such validation (Fig. 2.11) are presented for the situation on 24 April 2008 over a cloud system (with some precipitation in its middle) in USA (Fig. 2.12). Comparisons of finite differences (FD) and TL evolution for both the cloud radar reflectivity (Fig. 2.11 a,c) and the cloud lidar backscatter (Fig. 2.11 b,d) demonstrate that the TL model describing the evolution of finite-amplitude perturbations with the linearized observation operators generally fits well the finite differences between two NL integrations of observation operators. Larger differences are only observed in cloud radar reflectivity increments when there is a large amount of liquid water most probably due to non-linearities in the observation operator in such situations.
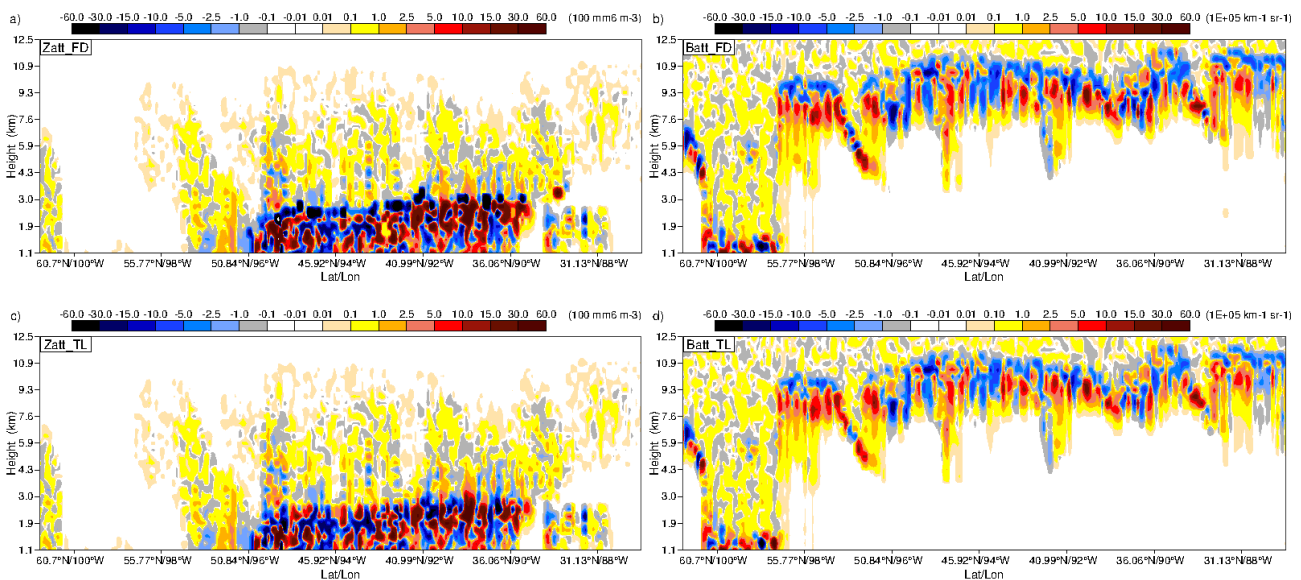

Figure 2.11: Increments of attenuated (a), (c) cloud radar reflectivity (in 100 mm^{-6} m^{-3}) and (b), (d) cloud lidar backscatter (in 10E+05 km^{-1} sr^{-1}) for the situation on 24 April 2008 over USA. (a), (b) finite-differences (FD) and (c), (d) tangent-linear (TL) model.
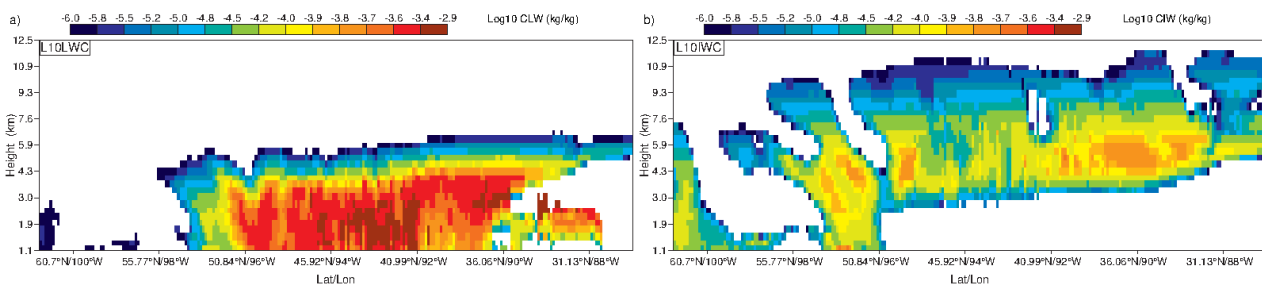

Figure 2.12: (a) Cloud liquid and (b) ice water content (in kg m^{-3} using logarithmic scale) for the situation on 24 April 2008 over USA.

2.5.2 *Adjoint version*

The adjoint of a linearized operator, $\mathbf{M}$, is the linear operator, $\mathbf{M}^*$, such that:

$$\forall \mathbf{x}, \forall \mathbf{y} \qquad < \mathbf{M}.\mathbf{x}, \mathbf{y} > = < \mathbf{x}, \mathbf{M}^*.\mathbf{y} > \tag{2.14}$$

where $<,>$ denotes the inner product and $\mathbf{x}$ and $\mathbf{y}$ are input vectors. The adjoint operator, for the simplest canonical scalar product $<,>$ (Eq. 2.14), is actually the transpose of the tangent linear operator, $\mathbf{M}^T$ (not its inverse).

For the practical verification of the adjoint code, one must test the identity described in Eq. (2.14). It is absolutely essential to ensure that the TL and AD codes verify Eq. (2.14) to the level of machine precision. Such verification was performed for both cloud radar and lidar observation operators and the required adjoint identity was achieved.

3   Pre-processing and handling: Scientific developments

This section describes the scientific developments related to the pre-processing and handling of cloud radar and lidar observations. The algorithms and definitions defined here are used to fill the Observation Data Base (ODB) definitions, which will be described in WP3000. All the developments have been for and tested with CloudSat and CALIPSO observations as they are extensive and readily available. The adaptations required for EarthCARE CPR and ATLID will be reviewed in WP4000.

IFS model setup for experiments

For all experiments in this section, model data values have been extracted from a set of 12 h forecasts generated using the ECMWF model with TCo639 spectral truncation (corresponding to approximately 18 km on a cubic octahedral grid) and 137 vertical levels. We use the CY43R1 version of the model, which was first operational in November 2016. The forecast results have been stored every half hour in order to compare observations in the same way as in the operational 4D-Var system, where all observations are split to half-hour time slots.

3.1   Quality control and screening

The aim of quality control and screening is to prevent or mitigate the influence of observations that will have a negative impact on the assimilation system. Screening ensures measurements are not assimilated where:

- they are unphysical,
- the forward model is not capable of representing the observations,
- there is an excessive non-linear relationship between the observation operator input and its output.

There are other benefits to screening. The total number of observations that can be assimilated is limited by time and computational constraints; screening reduces the volume of observations used. Initial screening is a balance between including as much information from observations as possible whilst preventing 'bad' observations from degrading the analysis.

Whereas screening blacklists observations for an objective reason, quality control reduces the weight of observations that are outliers relative to their expected error. All observations at ECMWF undergo quality control as a last step before assimilation. In this subsection we focus on the choice of initial screening indicators and thresholds.

Several indicators (see Table 3.1) have been chosen to screen observations where either the radar observation operator or the IFS itself are unlikely to represent reality. Observations are first screened by height. Although any ground return should be removed by the cloud mask during pre-processing, for safety we blacklist observations with 1 km of the surface. We also blacklist observations that are greater than 20 km above sea-level, where any signals are unlikely to be due to cloud. Next, CloudSat observations outside plausible bounds are removed. The minimum detectable signal is around -30 dBZ for CloudSat, and the maximum signal observed from hydrometeors is 20 dBZ due to a combination of Mie effects and strong attenuation in heavy precipitation (Kollias et al., 2007). To prevent any biases occurring from asymmetric sampling, the same thresholds are applied to the simulated reflectivities. Relaxing the minimum simulated reflectivity threshold might be beneficial for assimilation, but further tests need to be carried out.

Defining the thresholds for the remaining indicators in Table 3.1 is less straightforward. The blue line in Fig. 3.1a shows the percentage of observations that would pass screening for different maximum absolute

Indicator	Min	Max	Reason
Height (km)	1	20	Lower limit (relative to surface) to avoid surface return, upper limit (absolute) to discard spurious signals (although some stratospheric clouds may be removed)
CF_{IFS}	0.2	1.0	To avoid non-linearity and representativity issues
CF_{obs}	0.2	1.0	To avoid non-linearity and representativity
dBZ_{IFS}, dBZ_{obs}	−30	20.0	Plausible bounds for radar
$dB\beta_{IFS}, dB\beta_{obs}$	−50	0.0	Plausible bounds for lidar
FG departures	−20	20	Remove large departures
dBZ_{int}	0.0	41.3	Radar multpile scattering not modelled by observation operator
β_{int}	0.0	0.02	Avoid observations with excessive attenuation

Table 3.1: Screening thresholds for CloudSAT and Calipso observations.

first guess departures. In the STSE project, 20 dBZ (black dotted line) was found to be a good compromise between allowing the most observations and obtaining a successful minimization of the cost function in 1D-Var experiments. Using this threshold rejects less than 10% of observations.
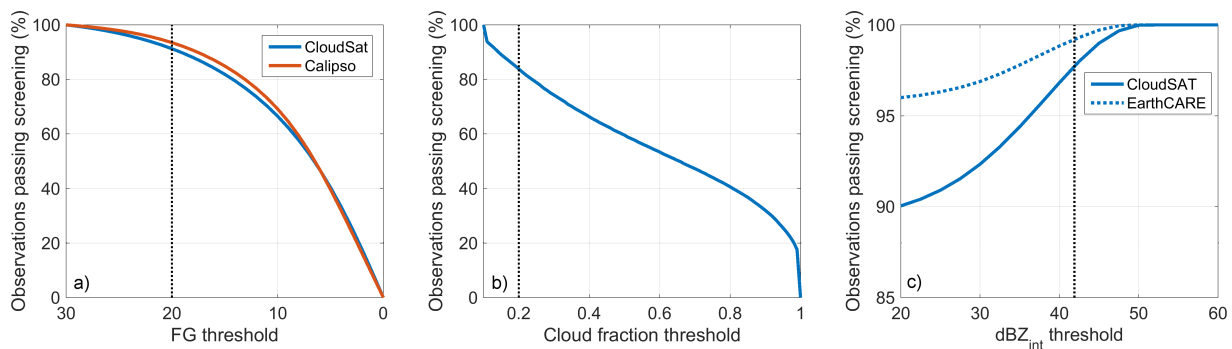

Figure 3.1: Percentage of observations passing screening for different thresholds of (a) first guess (FG) departures, (b) cloud fraction (defined as $\min\{CF_{IFS}, CF_{obs}\}$*), (c) integrated radar reflectivity. The dashed black line corresponds to the values given in Table 3.1.*

Cloud fraction is another potential indicator for screening. Both the model and observed cloud fraction could be used for screening. Observations where the superobbed cloud fraction is less than 1 are likely to fall into three regimes: near the edge of stratiform cloud; at the limit of the instrument sensitivity; or convective cloud (e.g., cumulus). Each of these options are likely to be difficult for the model to represent. Further, a severe mis-match between model and observed cloud fraction could indicate that the wrong regime is being modelled, which could be difficult for the assimilation system to correct.

Despite the complications of assimilating partially cloudy scenes, a compromise needs to be made to ensure the greatest information is extracted from the observations. Using a conservative cloud fraction threshold would discard many observations; as seen in Fig. 3.1b and Fig. 3.3a, a significant proportion (60%) of observations have either observed or modelled cloud fraction of less than 0.8. Whereas Fig. 3.1b uses a joint screening criteria, Fig. 3.3a differentiates between model and superobbed cloud fraction. From Fig. 3.3a we can see that a greater fraction of model data has CF less than 0.9, compared to the superobbed data, which tends to be more binary. Therefore, lowering the superobbed cloud fraction threshold will blacklist more observations than a matching reduction of model cloud fraction threshold. For now, a relaxed screening threshold of 0.2 is chosen, but this may need to be revised when assimilation experiments are performed.

To gain more insight into the effect of cloud fraction on combined model and observation errors, Fig. 3.3b and Fig. 3.3c show CloudSat radar reflectivity O-B statistics for different model and observed cloud fraction. As might be expected, observed radar reflectivity tends to be greater when the superob cloud fraction is greater. Similarly, the model radar reflectivity tends to be greater when the model cloud fraction is greater. There are differences between the two indicators when considering the O-B standard deviation; the error increases from around 8 dB for model cloud fraction greater than 0.8, to 12 dB for model cloud fraction around 0.2. Conversely, observed cloud fraction appears to have little effect on O-B standard deviation. One reason is that observations with cloud fraction less than one typically have a narrow range of values because the observations are often at the limit of radar sensitivity. A narrow range of observed values then reduces the chance of large FG departures, which in turn reduces the standard deviation of FG departures. These results are re-visited in Section 3.3, where, assuming the background error is relatively small, model cloud fraction can be used as a predictor of observation error.

The final screening task applied to radar reflectivity is to remove profiles where multiple scattering is suspected. Following Battaglia et al. (2011), the integrated reflectivity, dBZ_{int}, is used to blacklist profiles. Battaglia et al. (2011) showed that for CloudSat observations, integrating observations that exceed $8dBZ$ and using a threshold of 41.3 dBZ_{int} gave an optimum detection of cases where the magnitude of multiple scattering exceeds 3 dB. For EarthCARE's CPR, where multiple scattering is expected to be less due to a narrower field-of-view, they show the threshold for integrating observations increases to $12dBZ$. Applying the CloudSat thresholds to the CloudSat superobs results in the blacklisting of around 3% of observations (Fig. 3.1c), in agreement with Battaglia et al. (2011). To estimate the fraction of observations that will be screened with EarthCARE, if we apply the EarthCARE thresholds to the CloudSat data the percentage of observation blacklisted is around 1%.

For lidar observations, the final screening task is to remove observations that are likely to be severely attenuated. Attenuation can lead to ambiguities between clouds with a weak backscatter and clouds with a strong backscatter that is attenuated. As the actual attenuation in the observations is unknown, we use the integrated backscatter as a proxy for the amount of attenuation the signal has experienced. To prevent the introduction of biases, we can use the mean integrated backscatter of the actual and simulated observations. To choose the screening threshold in integrated backscatter, Fig. 3.2 shows the mean two-way transmission as a function of integrated backscatter using simulated observations. By setting the threshold in integrated backscatter at $0.02\,\mathrm{sr}^{-1}$, we will exclude observations that typically have a transmission of less than 0.1, which is equivalent to an attenuation of 10 dB.
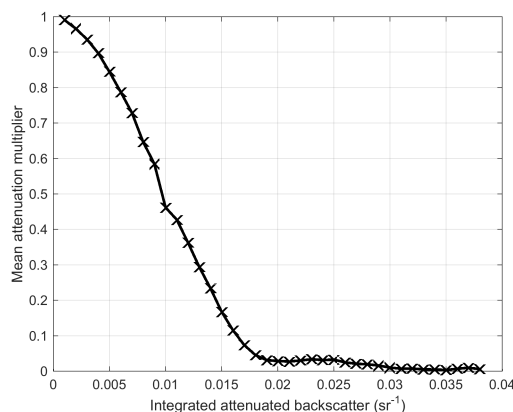

Figure 3.2: Mean two-way transmission as function of forward modelled integrated attenuated backscatter. For reference, a two-way transmission of 0.1 is equivalent to a 10 dB attenuation.
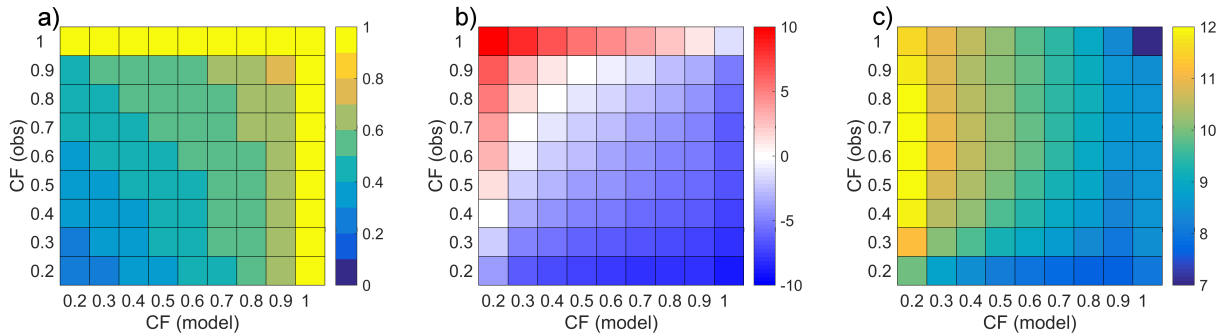
Figure 3.3: Effect of choosing different screening thresholds for model cloud fraction and observed cloud fraction. The 2D plots show (a) fraction of observations passing screening binned by cloud fraction threshold, (b) bias in FG departures binned by cloud fraction, (c) standard deviation in FG departures binned by cloud fraction. The plots were created using 31 days of CloudSat O-B statistics during August 2007.

Indicator	Range	Number of bins
Height	1 km - 20 km	21
Temperature	200 K - 300 K	42
Model hydrometeor type	6 different model species	6
Mean radar reflectivity	-30 dBZ - 20 dBZ	21
Mean lidar backscatter	-30 dBβ - 0 dBβ	13

Table 3.2: Bias correction indicators for cloud radar and lidar observations used in experiments. Bin edges are spaced linearly across indicator range.

3.2 Bias correction scheme

Bias correction may sound mundane, but it is an important component of the data assimilation system. While random errors are inherent in all observations and accounted for in the observation error matrix, systematic biases can have a detrimental impact on the analysis and must be removed a priori. For some well-understood observations, bias correction can be handled implicitly within the IFS 4D-Var system using Variational Bias Correction (VarBC; Dee, 2004; Aulign et al., 2007). For new observation types, such as those provided by EarthCARE, it is standard practice to apply a climatological bias correction scheme using long-term averages until the observation's biases and behaviour within the system are better known.

Regardless of whether the bias correction scheme is implicit or based on a climatology, its success depends on the choice of the indicators used. Indicators allow a different bias correction to be applied to the observations in different situations. At its simplest, a bias correction scheme might comprise a single global indicator to correct a fixed offset. This would be appropriate if, for example, it was known that the measurements contained were contaminated by a calibration error. In the case of satellite observations, while some calibration error is likely, the dominant source of bias is likely due to the microphysical assumptions made in the observation operators and will require several different indicators to achieve reliable performance.

So that different indicators can be tested and reviewed, a flexible and efficient off-line bias correction scheme has been implemented. Figure 3.4 details the steps taken for producing the bias correction tables. Firstly, historical 12 hr forecasts are generated as described in Sec. 3.1. Model equivalents are then found by running the updated version of ZmVar (Sec. 2). In parallel, CloudSat and CALIPSO observations are averaged to gridbox locations to create 'superobs'. These two pieces of information are combined to create First Guess (FG) departures. The observations undergo the same initial screening that is performed during assimilation so that the bias correction is not affected by observations that would be blacklisted operationally. This process is

repeated for different months to ascertain if there is any seasonality in the biases.
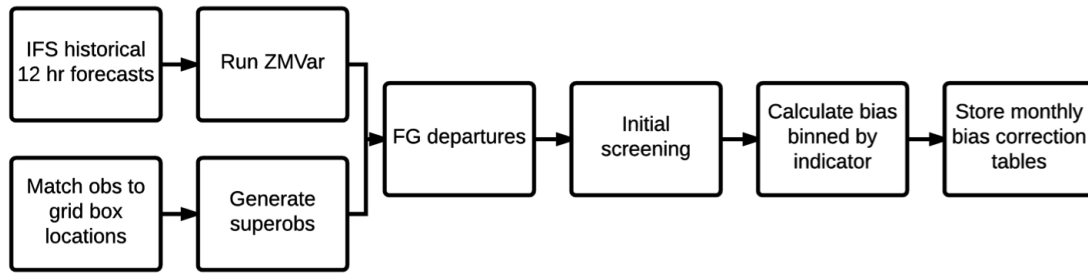

Figure 3.4: Flowchart showing the steps to obtain off-line bias correction look-up tables.

The bias correction scheme produces look-up tables for the climatological bias based on all or a subset of the shortlisted indicators listed in Table 3.2. In general, the more indicators that are used, the greater the apparent strength in performance of the bias correction scheme. However, it is possible to 'over-correct' the data; useful information contained within the observations could be lost if the bias correction is too aggressive. This is particularly the case if the choice of indicator is based on something that is solely based on either model or observation information. For example, if the model radar reflectivity was used as an indicator, regardless of the true bias, FG departures binned to lower model radar reflectivity values would appear to be too small. We now review the potential indicators before selecting a provisional subset to use for initial 4D-Var experiments.

Radar

Figure 3.5a shows the bias in radar reflectivity as a function of height for three different months. Below 10km, forward modelled radar reflectivity tends to be greater than the observations, whereas above 10km the observations tend to be greater. There is little seasonal variability in the bias below 12 km. Above 12 km there are some differences of up to 5 dB between months, however these values are likely to be affected by noise due to the smaller sample sizes; clouds that are detectable by CloudSat are rare above 14 km. Including height as an indicator is beneficial for two reasons; the first is that it could account for any systematic errors in our representation of the vertical propagation of the radar signal, and second that, coupled with temperature, provides a regime dependent indicator.

The bias in FG departures with respect to temperature (Fig. 3.5b) is remarkably uniform between 220 K to 290 K, where the model equivalent radar reflectivity around 2 dB greater than observations. There is a small jump in bias around 273 K, possibly due to the fact the different scattering properties of melting hydrometeors is not currently included in ZmVar's microphysical assumptions. Below 220 K the forward modelled radar reflectivity tends to be smaller than observations; it is difficult to diagnose whether this is as an issue with ZmVar's microphysical assumptions or an issue with the IFS moist processes without further investigation.

In addition to temperature, it would be tempting to choose the model water content as an indicator as it is a key input to the observation operators. However, as outlined above, using model-only (or observation-only) indicators is often sub-optimal and can introduce artificial biases. This is particularly true when considering indicators with a large uncertainty compared to the truth, such as model water content. Following Geer and Bauer (2011), one solution might be to use an indicator that is an average of model water content and a pseudo-retrieval of water content from observations. However, given that a retrieval of water content from either lidar backscatter or radar reflectivity would be highly uncertain, we choose to use the mean of the observation and simulated observations.
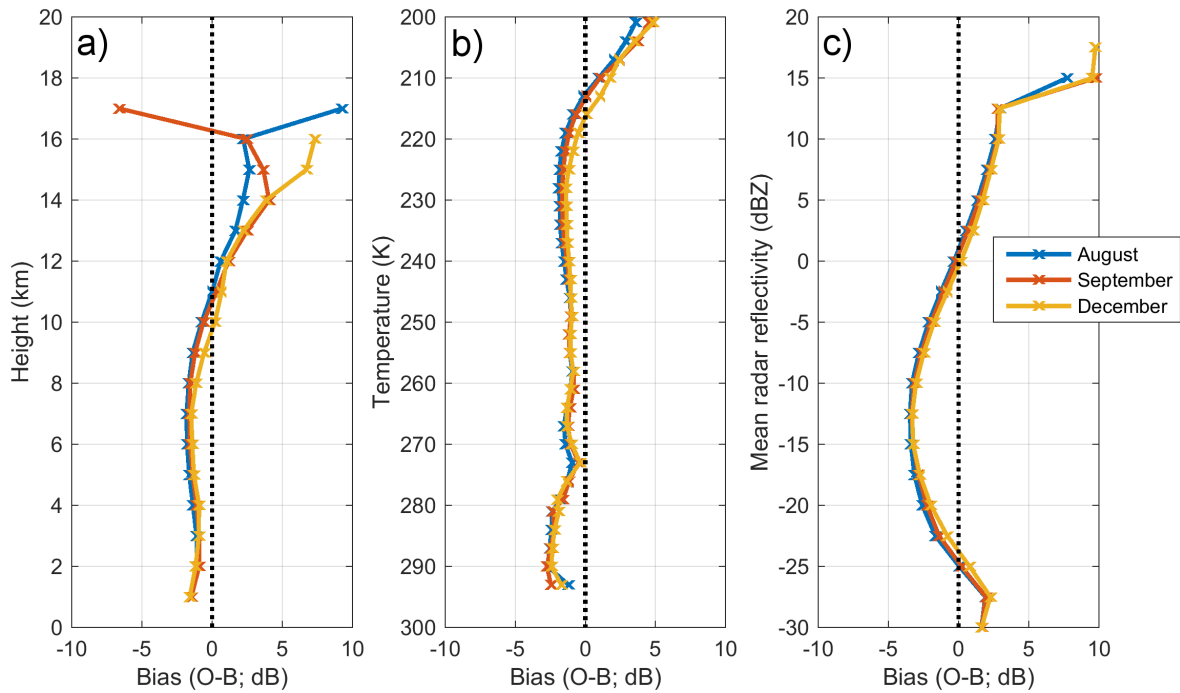
Figure 3.5: Bias correction for CloudSat observations as a function of (a) Height, (b) Temperature and (c) Mean radar reflectivity during 2007 for the months of August (blue line), September (red line) and December (yellow line).

For CloudSat, using the mean radar reflectivity between observations and model (Fig. 3.5c) again results in minimal seasonal variations, but there is a clear 'dipole' effect. For smaller mean radar reflectivity observations the model radar reflectivity tends to be greater than observations with a maximum bias around -15 dBZ. Conversely, for large mean radar reflectivity, the observations tend to be greater. This feature can also be seen in the Fig. 3.6a, where there is a cluster of observations around -28 dBZ with a model equivalent radar reflectivity of around -23 dBZ. The positive bias for larger mean radar reflectivity is also apparent between 0 and 10 dBZ, where observations are greater by 1-2 dB.
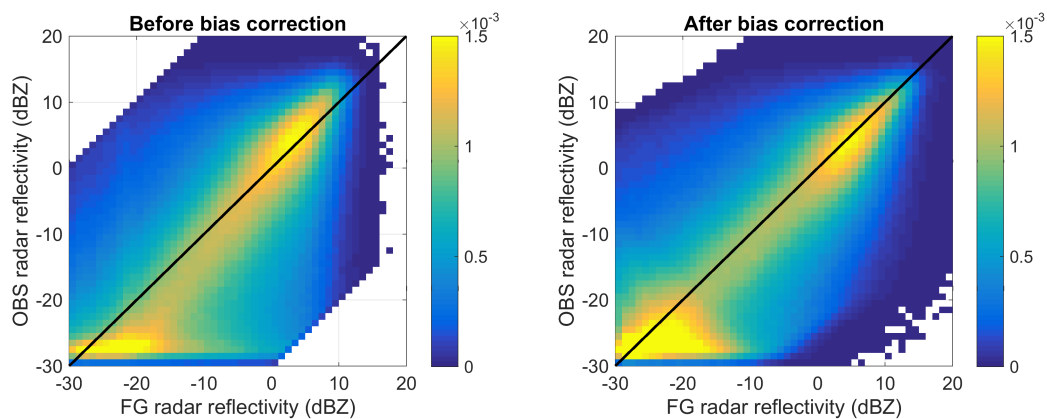

Figure 3.6: Joint probability density plots of simulated radar reflectivity and observed CloudSat radar reflectivity using observations during September 2007. The left panel shows data before bias correction, while the right panel shows the relationship after bias correction. Only data passing quality control are considered.

To further investigate, we use the 'dominant hydrometeor type' (defined as the model hydrometeor type with the greatest mass) as an indicator. By separating the first guess departures by hydrometeor type (Fig. 3.7), it becomes clear that the bias seen when observations are around -28 dBZ predominantly occurs when ice cloud is the dominant hydrometeor type. Further investigation (not shown) reveals that these points occur when the observed cloud fraction is less than unity. Given that the radar is close to the limit of its sensitivity (around -28 dBZ), many of these points could be false-positives. Assimilating observations that are at the limit of instrument sensitivity will require further testing during assimilation feasibility studies; it may be beneficial to increase the screening of observed radar reflectivity significantly above the limit of the radar sensitivity to reduce assimilation of false-positives. However, EarthCARE's increased sensitivity will allow the screening threshold to be lowered relative to CloudSat observations, allowing more observations to be assimilated.
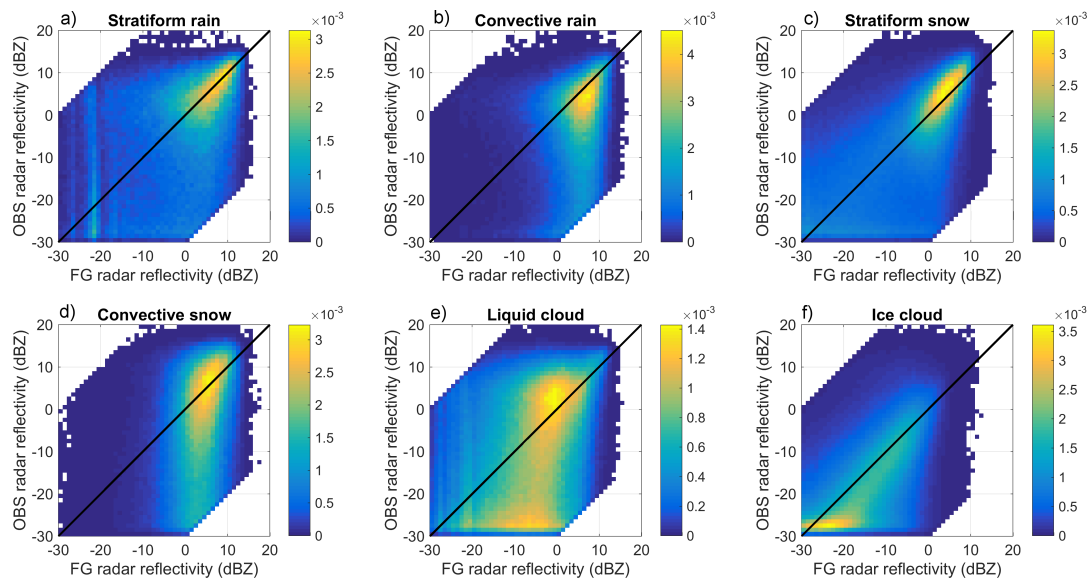

Figure 3.7: Joint probability density plots of simulated radar reflectivity and CloudSat radar reflectivity before bias correction during August 2007. Each panel shows the data for different dominant model hydrometeor types.

For the cluster of observations with positive O-B bias between 0 dBZ to 10 dBZ, it is clear from Fig. 3.7 that they occur in regions of stratiform or convective snow. In contrast, there is a negative O-B bias for convective rain, but stratiform snow dominates the probability density plots (Fig. 3.6) as it occurs more frequently. Both of these biases are probably due to the microphysical assumptions, but it is difficult to determine without comparisons with other observation types whether the bias is due to the particle size distribution or some other assumption, such as the assumed snow and ice particle shapes and their scattering properties. For now, the bias correction scheme ensures this bias is removed, but the microphysical assumptions should be revisited in the future to understand and potentially correct the cause of the bias at its source.

Applying these four indicators (height, temperature, mean attenuated radar reflectivity and dominant hydrometeor type) to the CloudSat data is visualised in Fig. 3.6b and in Fig. 3.8b. To ensure a fair test, the bias correction scheme is calibrated using data from August 2007, but applied to the first guess departures during September 2007. Qualitatively, the bias correction scheme is performing as expected; the data tends to be clustered closer to the one-to-one line in Fig. 3.6. By integrating the bias with height, Fig. 3.8 shows the spatial pattern of mean error before and after bias correction. Before bias correction, there is a negative bias everywhere except over part of the Western Pacific. After bias correction, the magnitude of the bias is reduced generally, and the spatial pattern of error reduces noise, with no obvious physical basis. Quantitatively, the global mean error before bias correction is -1.4 dB and -0.15 dB after.
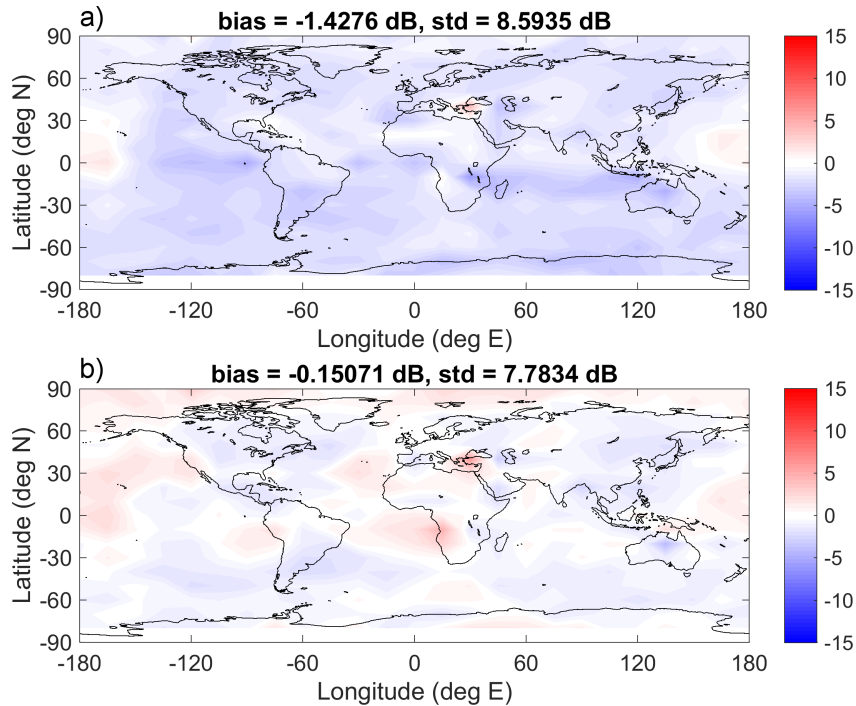
Figure 3.8: Spatial bias in radar reflectivity (a) before bias correction and (b) after bias correction during September 2007.

Lidar

A similar analysis of the bias correction scheme for lidar can be made using CALIPSO data. First, looking at the mean error with respect to height (Fig. 3.9a), between 4-14 km there is only a small mean error in first guess departures between CALIPSO and modelled attenuated backscatter. Below 4 km there is a tendency for observations to be greater, which increases towards the surface. Also, above 14 km observations tend to have a positive bias of 2-3 dB. While these errors are small, they hide compensating errors; Fig. 3.9b shows a complex pattern of error that is dependent on temperature. For temperatures between 210-235 K there is a negative bias, but elsewhere the bias is positive. The positive bias is greatest for temperatures above freezing, where it is typically 6-7 dB. For mean attenuated backscatter, the mean error is positive above -30 dBβ and negative below -30 dBβ.

To help explain these biases, Fig. 3.10a shows the joint density of first guess and observed backscatter. There are two clear density maxima; the first is around -20 dBβ and corresponds predominantly to observations within ice cloud. The second cluster of observations relates to first guesses just under -30 dBβ, where observations are between -40 and -15 dBβ. This second cluster corresponds mainly to stratiform and convective snow. While tempting to increase the backscatter of snow in the scattering look-up tables, the bias could be for other reasons, such as the attenuation of the lidar signal in the cloud above the snow or perhaps our parameterization of multiple scattering. Further research is required to pinpoint the source of these biases, perhaps using ground-based lidars that can 'see' snow without a-priori being attenuated by cloud.

Whereas mean radar reflectivity and dominant hydrometeor type were used as indicators for the radar bias correction scheme, for lidar, it is difficult to use indicators that could be pre-conditioned to a particular model state. For example, any modelled observation of snow or rain will already be attenuated, making a comparison with observed backscatter difficult, as it may not have undergone the same attenuation. For this reason, we only use height and temperature as indicators for the lidar bias correction scheme to avoid corrections that could be

pre-conditioned to the amount of attenuation in the modelled backscatter, rather than having a physical basis.

Despite only using two indicators, the lidar bias correction scheme performs well. As for the radar case, to ensure a fair test, the bias correction scheme is calibrated using data from August 2007, but applied to the first guess departures during September 2007. Qualitatively, in Fig. 3.10, the observations and first guesses are brought closer to the one-to-one line. The global mean bias is reduced from 0.32 dB to 0.11 dB, although some of the spatial pattern is retained (Fig. 3.11).
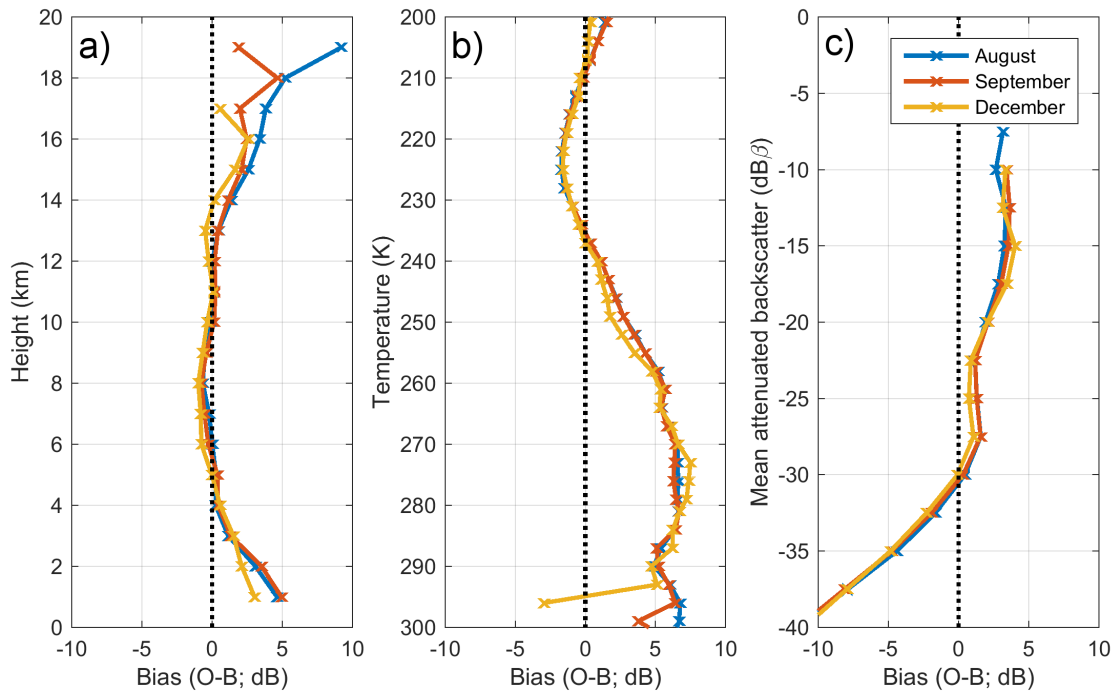

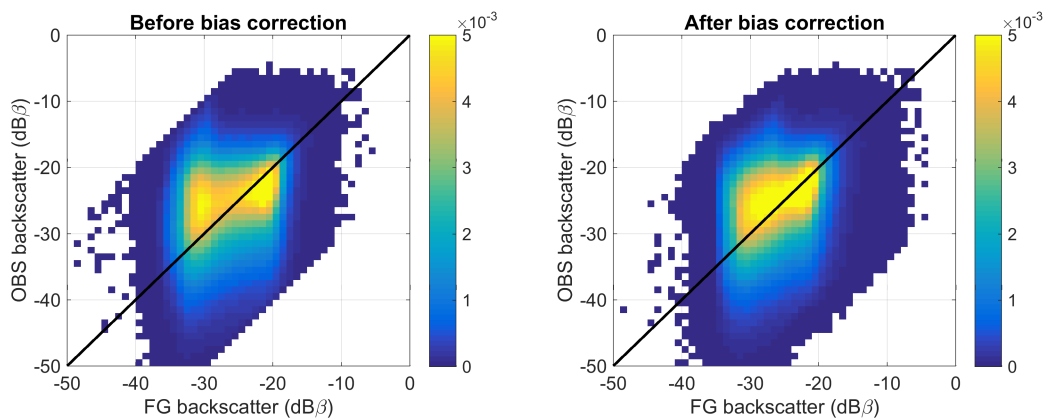Figure 3.9: Same as Fig. 3.5, but for CALIPSO observations.

Figure 3.10: Same as Fig. 3.6, but for CALIPSO lidar attenuated backscatter observations.
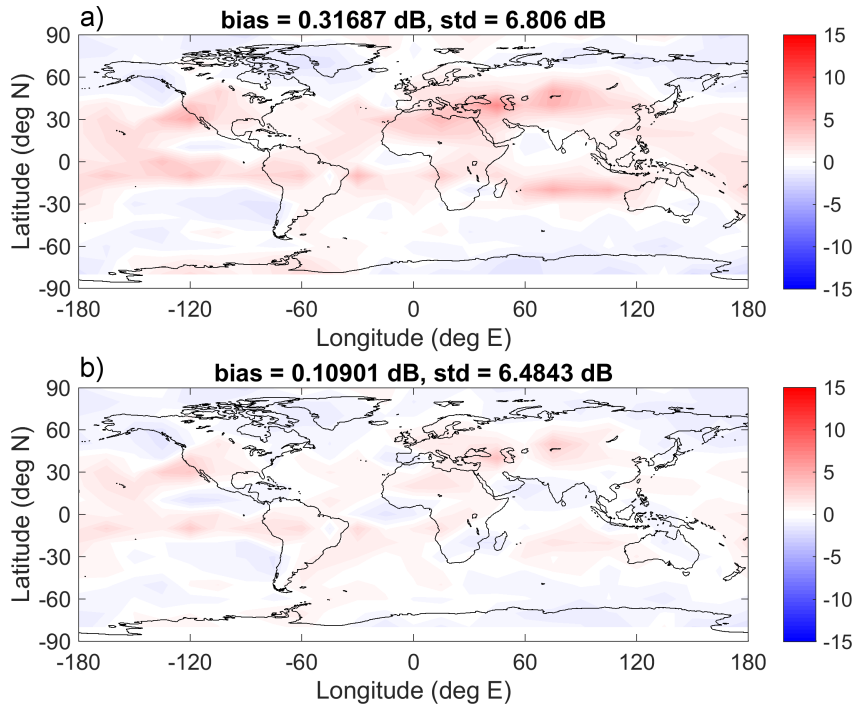
Figure 3.11: Same as Fig 3.8, but for CALIPSO lidar attenuated backscatter observations.

3.3 Observation error definition

The aim of a variational data assimilation system, such as 4D-Var, is to find an optimal model state that minimizes, in a least-square sense, the distance to both the observations $\mathbf{y}^o$, and to an *a-priori* model state $\mathbf{x}^b$ called the background. The misfit is measured by the following objective cost-function $J(\mathbf{x})$, to be minimized during the assimilation process:

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})), \qquad (3.1)$$

where H is the observation operator, which provides the model equivalents to the observations. In our case, for radar and lidar observations, the contribution to H is given by ZmVar. $\mathbf{B}$ is the background error covariance matrix; as well as providing the uncertainty in the *a-priori* model state, crucially, its correlations allow observations to influence the model state away from the observations' immediate locations. Finally, $\mathbf{R}$ is the observation error covariance matrix. In conjuction with $\mathbf{B}$, $\mathbf{R}$ determines the weight given to each observation during the assimilation process. At ECMWF, $\mathbf{B}$ is specified by a combination of climatological statistics and the Ensemble of Data Assimilations (EDA; Isaksen et al., 2010). Unlike the background error covariance, the observation error must be specified for each observation.

There are many different methods to estimate $\mathbf{R}$. When $\mathbf{R}$ is expected to be small, for example for conventional observations of temperature or humidity, a fixed value is often used. Other methods exploit statistical relationships between the model variables and observations, (e.g., Hollingsworth and Lönnberg, 1986; Desroziers et al., 2005). In a related approach, for microwave radiance observations, (Geer and Bauer, 2011), show that using a climatology of first guess departures binned by cloud amount is a significant improvement over using fixed values. Finally another approach is to separate the error into its constituent parts and define the error from first principles.

We choose the latter approach, defining the observation error explicitly based on physical understanding for

two reasons. The first is that, owing to the profiling nature of the observations, the true observation error is likely to be highly situation dependent; for example, the error at a particular model level will depend on the optical depth of the layer above. Secondly, at the time EarthCARE becomes operational, we will not have a long history of observations to generate a climatological observation error covariance matrix.

Under the hypothesis of uncorrelated errors, $\mathbf{R}$ can be written as:

$$\mathbf{R} = \mathbf{E} + \mathbf{F} + \mathbf{O}, \tag{3.2}$$

where $\mathbf{E}$ the instrument error, $\mathbf{F}$ the observation operator error, and $\mathbf{O}$ the representativity error. In the case of measurements from profiling instruments, the off-diagonal elements of each of these matrices represent the covariance across measurements in the horizontal and vertical. In this work, this correlation is neglected, therefore we will use the term observation error to refer to the diagonal of each error covariance matrix. We will now review and define the errors for each component of $\mathbf{R}$.

3.3.1   Instrument error

Radar

The instrument error is the random error in the measurement due to noise. For cloud radars, the uncertainty in radar reflectivity in logarithmic units, ΔZ_{dB}, can be estimated using the following equation:

$$\Delta Z_{dB} = \frac{4.343}{\sqrt{N}} (1 + \frac{1}{SNR}), \tag{3.3}$$

where SNR is the linear signal-to-noise ratio and N is the number of independent samples. For CloudSat and the EarthCARE CPR, the motion of the satellite relative to the instrument's footprint means each pulse is independent and therefore N is equal to the number of samples.

To compute the SNR, an estimate of the instrument noise is required. For testing using CloudSat data we follow Delanoë and Hogan (2010) and approximate the radar background noise, N_{dBZ} as:

$$N_{dBZ} = -131.4 + 20 log_{10}(H_{sat}), \tag{3.4}$$

where H_{sat} is the distance from the satellite to the target volume in metres. For the EarthCARE CPR a parameterization such as Eq. 3.4 will not be possible until after launch. Therefore, we will follow an alternative approach (Hildebrand and Sekhon, 1974) that is implemented in the L2 C-PRO algorithm:

1. For each profile, the power, $P(i)$, is sorted in ascending order
2. Set n=N, where N is the total number of range gates
3. While the profile is determined to be coloured noise set n=n-1, i.e.,

$$n \sum_{i=1}^{n} P(i)^2 - 2 \left(\sum_{i=1}^{n} P(i) \right)^2 = \begin{cases} < 0, & \text{White noise} \\ > 0, & \text{Coloured noise} \end{cases} \tag{3.5}$$

4. When the profile is determined to be free from signal (white noise), the mean noise power is calculated by averaging the remaining gates.

Further details are provided in the L2 C-PRO documentation (Kollias et al., 2016).

Lidar

The uncertainty in lidar backscatter for the EarthCARE ATLID is expected to be provided in the L1B product. For testing using CALIPSO data, the measurement uncertainty needs to be specified. Following the work of Liu et al. (2006), the standard deviation $\Delta\beta$ of random errors (due to shot noise) in the measured lidar backscatter β can be expressed as:

$$\Delta\beta = \{NSF^2\beta + \left(\frac{r^2}{C}\right)^2 \left[(\Delta V_b)^2 + (\Delta\bar{V}_b)^2\right]\}^{\frac{1}{2}} \tag{3.6}$$

where r is the distance (in meters) of each lidar gate from the satellite and C is the lidar calibration constant such that the signal power V is: $V = C\beta r^{-2}$. NSF is the Noise Scale Factor representing the effect of the photomultiplier tube to increase the noise above what would be expected purely from Poisson statistics. All these quantities are included in CALIPSO Level 1B product.

ΔV_b in Eq. 3.6 is the standard deviation of the background signal power. It can be evaluated using (N) measurements only affected by molecular backscatter (e.g. using the ones above 30 km). $\Delta\bar{V}_b$ is the standard deviation of the mean background signal and therefore: $\Delta\bar{V}_b = \Delta V_b/\sqrt{N}$. It should be noted that when the measured backscatter values are obtained as an average of n samples onto a lower resolution grid, the standard deviation of random errors for the averaged measurements $\Delta\bar{\beta}$ is obtained from the original ones $\Delta\beta$ as:

$$\bar{\Delta\beta} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\Delta\beta_i^2} \tag{3.7}$$

3.3.2   Observation operator error

To convert model hydrometeor content into either radar reflectivity or lidar backscatter, many different assumptions are made, each with the potential to introduce error to the forward modelled observations. While the bias correction scheme should account for any systematic error present, random error remains and must be accounted for each observation point. Possible sources of error include:

- Radiative transfer in scattering models
- Hydrometeor shape
- Particle size distribution
- Particle density
- Multiple scattering
- Subgrid assumptions (including overlap, inhomogeneity and convective precipitation fraction)

The first four sources of error relate to converting a given model hydrometeor water content into bulk single scattering properties and are therefore only dependent on the wavelength of the instrument. Neglecting error due to the radiative transfer in scattering models (negligible for Mie, typically less than 5% for ray tracing or the discrete dipole approximation), we shall refer to this error as 'microphysical error', which, in part, originates from the model's incapability of resolving the regional microphysical variability. The remaining sources of error are dependent on both the model configuration and the instrument itself. For example, errors in multiple scattering depend on how/if the observation operator simulates it and the field-of-view (FOV) of the instrument, while errors in subgrid assumptions predominantly depend on the size of the model gridbox. We shall refer to this second group of error as 'driver error'.

The challenge of accounting for microphysical error is universal and its application is not restricted to data assimilation; Kulie et al. (2010) attempt to quantify its magnitude for remote sensing. Driver error can be

Hydrometeor type	Parameter and mean value	Perturbation (%)
Cloud liquid	$N_t = 100\,\mathrm{cm}^{-3}$	20
	$\sigma_g = 0.3$	20
Cloud ice	$a = 0.0094$	10
	$b = -0.87$	10
Rain LS	$x_1 = 0.22$	20
	$x_2 = 2.2$	0
Snow LS	$a = 0.0026$	10
	$b = -1.42$	10
Rain conv.	$N_L = 0.08\,\mathrm{cm}^{-4}$	50
	$\mu = 5.0$	30
Snow conv.	$a = 0.0026$	10
	$b = -1.42$	10

Table 3.3: List of parameters that are perturbed to obtain the component of observation operator error due to uncertainty in the particle size distribution.

quantified by comparing a more complex or 'perfect' driver that makes less assumptions or is more physical (such as the 'complex' version of zmvar, e.g. Di Michele et al., 2012).

To model the microphysical error specific to ZmVar, we take a Monte Carlo approach by perturbing various parameters of the microphysical assumptions used to generate the bulk scattering properties (see Table 2.1) within physical bounds. Table 3.3 lists the parameters of the particle size distributions that are perturbed for the six hydrometeor types used within the IFS. Random Gaussian noise is added to the mean value for each perturbed parameter with standard deviation equal to the size of the perturbation listed in Table 3.3. In addition to error from the PSD, the uncertainty in particle shape for solid hydrometeors is modelled by randomly selecting different particle shapes; for lidar, one of eight particles are chosen from Yang et al. (2000), for radar one of 11 are chosen from Liu (2008). This process is repeated until 100 different, physically plausible realisations of the scattering look up table are generated. The uncertainty in the bulk scattering properties is then calculated by finding the standard deviation as a function of water content and temperature.

Figure 3.12a shows the error in single scattering radar reflectivity as a function of water content. The uncertainty in liquid cloud is found by perturbing the droplet number concentration and size parameter of the lognormal size distribution. The magnitude of the perturbation is guided by variability in number concentration and shape parameter of in situ observations (e.g., Miles et al., 2000). The uncertainty is invariant with changes in water content for two reasons; the shape of the size distribution is independent of water content by definition, and also the cloud droplets are always small enough to be firmly in the Rayleigh scattering regime.

Choosing how much to perturb the parameters for the rain PSD was not as straightforward as with liquid cloud as the PSD is based on an empirical fit to measurements, which cannot be directly observed. However, using Figure 3 from Abel and Boutle (2012) as a guide, we choose to perturb x_1 by 20%, which spans the range of observed λ and N_0 without needing to perturb parameter x_2 (see Table 2.1 for parameter definitions). Interestingly, the resulting uncertainty in radar reflectivity (Fig. 3.12a) has a strong dependence on water content; for larger water content, the uncertainty tends to be smaller because the size distribution includes drops that undergo Mie scattering that reduces the sensitivity to changes in drop size relative to drops that are in the Rayleigh scattering regime.

For both ice and snow, ZmVar assumes a particle size distribution given by Field et al. (2007), which is tuned to aircraft measurements and is a function of temperature, water content and the assumed particle density. Again, as for the parameterization of rain, it is not obvious how to perturb the parameters in the empirical fits in a
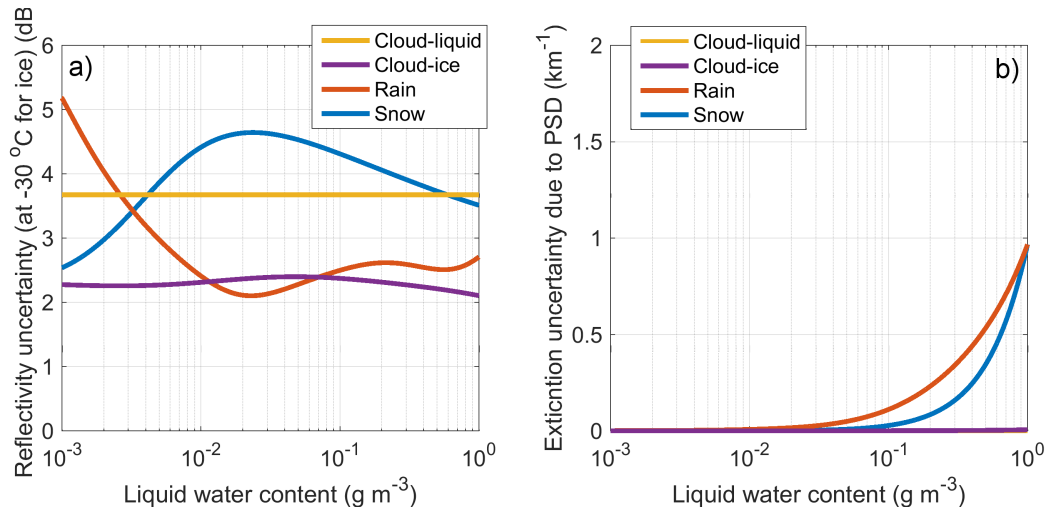
Figure 3.12: Uncertainty $(1-\sigma)$ in (a) radar reflectivity (dB) and (b) extinction as a function of water content for a fixed temperature of $10\,°C$ for liquid hydrometeors and $-30\,°C$ for solid hydrometeors at 94 GHz. The uncertainty is found by perturbing parameters of each hydrometeor type's particle size distribution.

physical way. However, the particle size distribution is also a function of the mass-size relationship:

$$m = aD^b \tag{3.8}$$

where m is the mass, D is the diameter and b is the fractal dimension. In the literature, b ranges from between 1.9-2.1 for snow aggregates (e.g., Locatelli and Hobbs, 1974; Mitchell, 1996; Heymsfield et al., 2002). After converting between the density-size relationship, $(\rho(D) = a_d D^{b_d})$ and the mass-size relationship using:

$$a_d = \frac{6}{\pi} a 10^{-(3+2b)} \tag{3.9a}$$

$$b_d = b - 3. \tag{3.9b}$$

we choose to perturb a_d and b_d by 10%. The idea is that by perturbing the mass-size relationship, we also implicitly account for variability in the PSD. To account for variability in the ice habit and corresponding single particle scattering properties, for each realisation of the look-up table, we randomly select one of the eight ice particle habits available in the Yang database. The resulting uncertainty in radar reflectivity as a function of water content is shown by the blue line in Fig. 3.12a. As water content increases, the uncertainty also increases as the PSD tends to broaden for larger water content. After a peak at around $0.02\,\mathrm{g\,m^{-3}}$, the uncertainty decreases due to the transition from the Rayleigh to the less-sensitive Mie scattering regimes.

For ice cloud, the same approach is taken as for snow, except that the temperature within the Field et al. (2007) parameterization is fixed at 203 K. At this low temperature, the PSD is narrower and the mean diameter is smaller, resulting in a reduced uncertainty compared to snow, with less dependence on water content. An uncertainty of around 2 dB is consistent with aircraft observations of ice cloud Hogan et al. (2006).

The uncertainty in extinction ($\Delta\sigma_{ext}$) for the different hydrometeor types is shown in Fig. 3.12b. In the Rayleigh scattering regime, extinction is approximately proportional to the water content, so the uncertainty is zero for liquid and ice cloud, where the particle sizes are small relative to the radar wavelength. The extinction uncertainty grows for rain and snow water contents when the PSDs contain sizes large enough so that the extinction is sensitive to particle size and shape.

Figure 3.13 shows the uncertainty in lidar backscatter using the same approach and perturbations (Table 3.3) as for the uncertainty in radar backscatter. The uncertainty in ice and snow particle habit is modelled by randomly selecting one of the eight habits available in Yang et al. (2000). Uncertainties are greatest for precipitating hydrometeor types, where the PSD contains a greater range in particle sizes. For rain, the greatest uncertainty can be for smaller water contents, where the lidar backscatter is sensitive to the number of smaller, and collectively more reflective, drops.
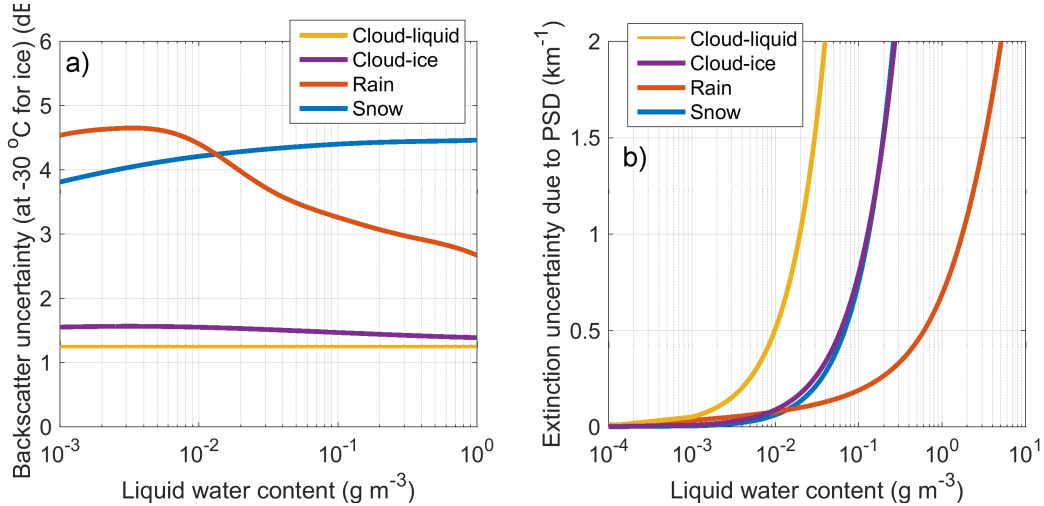

Figure 3.13: Same as Fig. 3.12, but for lidar backscatter at 532 nm.

The extinction uncertainty for lidar (shown in Fig. 3.13b) is greater for lidar backscatter than for radar reflectivity. The greatest uncertainty is seen for liquid cloud due to the greater extinction coefficient associated with small cloud droplets. Rain, which typically has drops with large diameter, provides the least uncertainty in extinction coefficient, however, for space-based lidar, any observations of rain are likely to already be highly uncertain due to the strong likelihood of an optically thick layer of cloud above.

The total observation error accounting for errors in both scattering and absorption is then calculated as (here shown for attenuated radar reflectivity):

$$\Delta Z_{dB} = 4.343 \sqrt{(\Delta \ln Z)^2 + 4 (\Delta \tau)^2}. \tag{3.10}$$

Assuming the unresolved microphysical variability error is uncorrelated, the uncertainty in optical depth is:

$$\Delta \tau = \sqrt{\sum_{l=1}^{n_l} dz_l \Delta \sigma_{ext}^2}, \tag{3.11}$$

where n_l is the number of levels above the lth model level. An analogous expression is used for estimating the uncertainty in attenuated lidar backscatter. Currently, random errors due to driver error are assumed to be negligible relative to the microphysical error.

3.3.3 Representativity error

It would be almost impossible to devise a system capable of observing the exact contents of one model gridbox during one model timestep. Therefore, all observations contain some form of representativity error, which describes the degree of mis-match between the observation scale and model scale in both space and time. Representativity error is deemed to be small compared to other errors for many conventional observations (e.g.,

radiosondes; Kitchen, 1989). In the case of satellite observations of cloud using radar and lidar with narrow FOVs, the representativity (or lack of) of the observations is a significant problem and, as we will see, dominates the observation error in the majority of cases. In the succeeding discussion, for simplicity, we will neglect the temporal representativity error problem. Regardless of the model timestep, within the IFS 4D-Var assimilation system each observation time is quantised to 30 minute intervals, so we assume clouds to remain fixed within each gridbox during this time.

In recognition of the dominance of representativity error to profiling observations from radar and lidar, significant effort was allocated to developing the Structure Function Maximum (SFM; Stiller, 2010) method to quantify representativity errors during the previous QuARL project (Janisková et al., 2010). Although the method shows promising results when validated using TRMM rain rates and MODIS cloud fractions, it is heuristic in nature and relies on the validity of synthetic data created using climatological statistics.

Here we propose a simpler method that is based upon the assumption that the local variability of measurements along the satellite track is representative of the gridbox variability, and that the spatial correlation between measurements can be approximated using a climatological correlation. We will first define the method and its mathematical basis, then validate the technique using well-behaved synthetic data and then more challenging 'real' data using MODIS optical depth retrievals. Next we show how the method can be applied to CloudSat and CALIPSO data and detail a simplification for operational data assimilation.

Method

The basis of the method is to treat the representativity error as a sampling error. Formally, let us define each observation as q_i at position i along the track. Our estimate of the gridbox mean, $\bar{q}$, is simply:

$$\bar{q} = \frac{1}{n} \sum_{i=1}^{n} q_i, \tag{3.12}$$

where n is the number of observations that fall within the gridbox. If the q_i are not correlated, then the sampling error in the estimate of the true mean is:

$$\mathbf{E}\left\{ (\bar{q} - \mu)^2 \right\} = \frac{\sigma^2}{n}, \tag{3.13}$$

where μ is the true mean (i.e. the mean of observations if the whole gridbox was sampled) and σ is the standard deviation of q. Again, if the q_i are not correlated, then, S^2, the sample variance, given by:

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (q_i - \bar{q})^2 \tag{3.14}$$

is an unbiased estimate of σ^2.

However, clouds and cloud-related measurements are highly correlated, which leads to biases in both Eq. 3.13 and Eq. 3.14. Anecdotally we know clouds are correlated, if you were to look up at a cloud through a pinhole, chances are that there would be cloud surrounding your view, and not a random chance of cloudy and cloud-free areas. More rigorously, cloudy observations of radar reflectivity have been known to be correlated for some time (e.g., Heymsfield, 1976; Smythe and Zrnic, 1983). Fortunately, if the correlation of x_i is known then it can be shown (see Anderson, 1994) that:

$$\mathbf{E}\left\{ (\bar{q} - \mu)^2 \right\} = \frac{\sigma^2}{n} \left[1 + 2 \sum_{k=1}^{n} \frac{(n-k)}{n} \rho(k) \right] \equiv \frac{\sigma^2}{n} \gamma_2, \tag{3.15}$$

where ρ is the correlation in q_i. Similarly, the estimate of the population standard deviation using the sample standard deviation can be corrected:

$$\mathbf{E}\left\{S^2\right\} = \sigma^2\left[1 - \frac{2}{(n-1)}\sum_{k=1}^{n}\frac{(n-k)}{n}\rho(k)\right] \equiv \sigma^2\gamma_1. \tag{3.16}$$

Substituting Eq. 3.16 into 3.15 leads to:

$$\mathrm{Var}(\bar{q}) = \left(\frac{\mathbf{E}\left\{S^2\right\}}{n}\right)\frac{\gamma_2}{\gamma_1}, \tag{3.17}$$

which provides an estimate of the sampling error in $\bar{q}$. Although this method shows how it is possible to find an analytical solution to correcting the sampling error for correlations, it has several shortcomings. Firstly, the method knows nothing about the geometry or area of the gridbox; in fact it assumes the gridbox is effectively 1D (or the width of the instrument FOV in one dimension and infinitely long in the other). It also assumes that the standard deviation of the gridbox is equal to the global standard deviation.

To account for the finite size of the gridbox and its 2D nature, let us define a new statement of the problem, this time in continuous space and assuming a rectangular gridbox:

$$\bar{q}^{1D} = \frac{1}{H_x}\int_0^{H_x} q(x,y_0)dx \tag{3.18a}$$

$$\bar{q}^{2D} = \frac{1}{H_xH_y}\int_0^{H_x}\int_0^{H_y} q(x,y)dxdy, \tag{3.18b}$$

where $\bar{q}^{1D}$ is the average of measurements along a transect of length H_x at position y_0 and $\bar{q}^{2D}$ is the average of observations if the entire gridbox was sampled. We would like to estimate the mean square error between $\bar{q}^{1D}$ and $\bar{q}^{2D}$, i.e.:

$$\begin{aligned}(\bar{q}^{1D} - \bar{q}^{2D}) &= \frac{1}{H_x^2H_y^2}\left(\int_0^{H_x}\int_0^{H_y} q(x,y_0) - q(x,y)\right)^2 \\ &= \frac{1}{H_x^2H_y^2}\int_0^{H_x}\int_0^{H_y}\int_0^{H_x}\int_0^{H_y}(\Delta q(x_1,y_1)\Delta q(x_2,y_2))dx_1Dy_1Dx_2Dy_2,\end{aligned} \tag{3.19}$$

where

$$\Delta q(x_1,y_1)\Delta q(x_2,y_2) \equiv [q(x_1,y_0) - q(x_1,y_1)][q(x_2,y_0) - q(x_2,y_2)] \tag{3.20}$$

By definition, the expectation of Eq. 3.20 is given by:

$$\begin{aligned}\langle\Delta q(x_1,y_1)\Delta q(x_2,y_2)\rangle = \sigma^2\Big[&\left[\rho(\|\mathbf{x_{10}} - \mathbf{x_{20}}\|) + \rho(\|\mathbf{x_{11}} - \mathbf{x_{22}}\|)\right] - \\ &\left[\rho(\|\mathbf{x_{11}} - \mathbf{x_{20}}\|) + \rho(\|\mathbf{x_{10}} - \mathbf{x_{22}}\|)\right]\Big]\end{aligned} \tag{3.21}$$

where $\|...\|$ is the Euclidean norm and

$$\begin{aligned}\mathbf{x_{i0}} &= \begin{pmatrix} x_i \\ y_0 \end{pmatrix} \quad \text{for } i = 1,2 \\[6pt] \mathbf{x_{ii}} &= \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{for } i = 1,2\end{aligned} \tag{3.22}$$

.

By substituting eq. 3.21 into eq. 3.19, the expected sampling error of a 1D measurement within a 2D gridbox can be found. Equation 3.19 can then be solved numerically by discretising the integrals to the measurement resolution. All that is required is an estimate of the population standard deviation, σ, and the correlation function, ρ. The estimate of the population standard deviation can be found locally by taking the sample standard deviation along the measurement track and inverting eq. 3.16. Assuming that a fixed global correlation function is obtained, and the gridbox and measurement track are fixed, the representativity error can be found by multiplying the local standard deviation of measurements along the track by a single, 'scale factor'. The procedure of computing the scale factor for cloud radar and lidar measurements within a data assimilation system is discussed later.

Validation using synthetic data and MODIS optical depth data

To evaluate the new methods, we can test the methods with 2D fields where the true representativity error is known. We first test the methods using synthetic data with a prescribed log-normal distribution and correlation (Fig. 3.14a). The synthetic data is generated using the same method as Stiller (2010), which was based on the technique of Pardo-Igúzquiza and Chica-Olmo (1993, 1994). The correlation function is chosen to be similar to the correlation function seen in MODIS optical depth retrieval scenes (e.g., Fig.3.14b). The 2D scene is then subset into 20 km x 20 km 'gridboxes', and a 'measurement track' is simulated across the centre of the gridbox. The true representativity error for the scene is then the average difference between the mean of the measurement track and the mean of the gridbox.
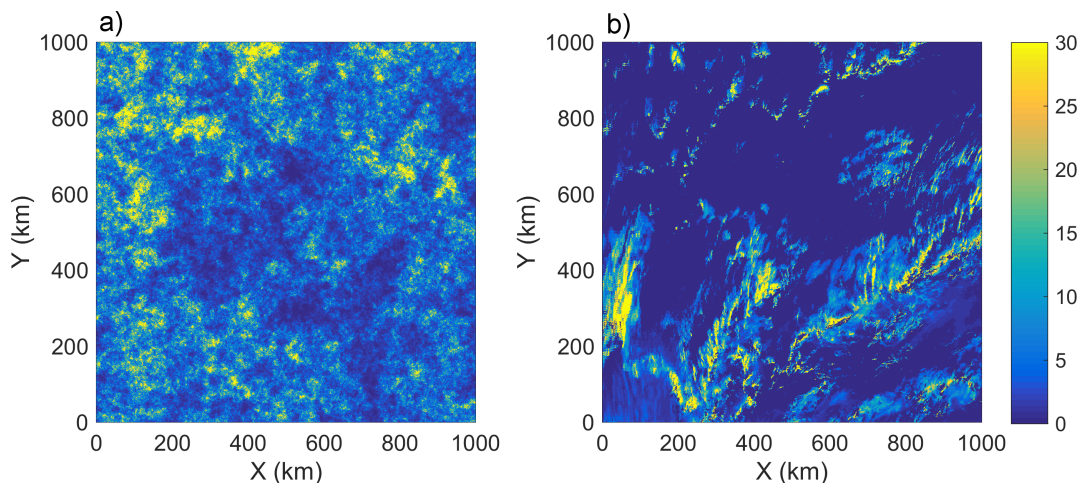
Figure 3.14: Examples of the 2D data used to evaluate the new representativity error method. (a) shows synthetic data with prescribed statistical properties and (b) a MODIS optical depth scene.

Using this approach, we test both the '1D method' (eq. 3.17) and the '2D method' (eq. 3.19). Using the 1D method (blue lines in Fig. 3.15), it is clear that the mean errors are overestimated across all magnitudes of error. This is to be expected; the 1D method does not account for the reduced variability within the finite gridbox due to the correlations in the measurements, leading to an over-estimation of the error. Using the 2D method (red lines) corrects for this error and predicts the actual representativity error almost perfectly.

While the 2D method predicts the representativity error extremely well using the synthetic data, 'real' data is unlikely to have consistent statistical properties across the domain. To account for this, we also test the methods on a variety of MODIS optical depth retrieval scenes. To ensure a fair test, we obtain the 'global' correlation function from one MODIS scene and then assume that this correlation function is representative of the other scenes. A total of 15 different scenes, randomly chosen from all the scenes available for August 1st 2007, are

tested.

Again, the 1D method tends to overestimate the representativity error (Fig. 3.15), although regions of the MODIS scenes with reduced variability are better estimated. The 2D method tends to underestimate the error, particularly when the predicted error is relatively small. It is probably to be expected that the 2D method will occasionally underestimate the representativity error, because in real data it is likely that there will be times when the local variability along the track is not representative of the grid box standard deviation. This could occur, for example, at the edges of clouds or where the scene is not completely overcast. Despite this the method clearly offers skill at predicting the representativity error, and is superior to the 1D method.
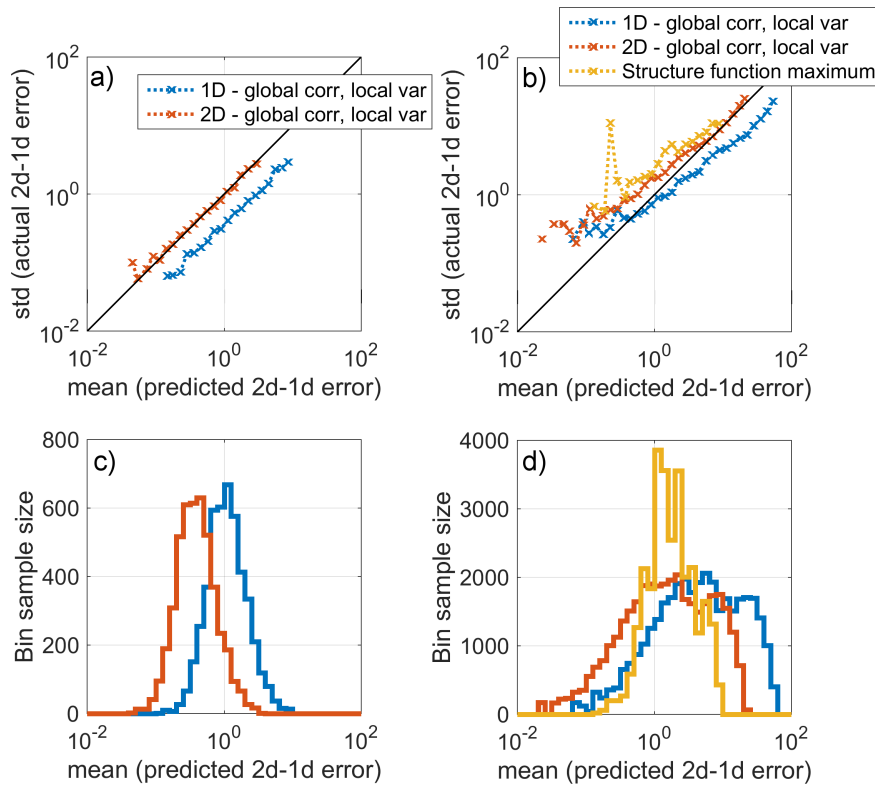

Figure 3.15: An evaluation of methods for estimating representativity error. (a) shows the standard deviation of actual error binned by the predicted error of the 1D (blue) and 2D (red) methods using synthetic data. (b) same as (a) but using 15 MODIS optical depth scenes and also shows the structure function maximum method (yellow). (c) sample size in each bin and method of (a), (d) same as (c) but for the MODIS scenes.

Finally, we also compare the methods to the structure function maximum (SFM) method. For fairness, the SFM method is tuned using the same MODIS scene that was used to obtain the correlation function for the other methods. For the MODIS scenes, the 2D method tends to perform better than the SFM method, with the SFM method tending to underestimate the representativity error. This could be partly due to the fact that the scene used to train the SFM method did not contain the full range of values of the structure function seen in all the scenes, limiting the methods ability to predict errors greater than 10 and smaller than 0.1 (see yellow line in Fig. 3.15).

Application to CloudSat and CALIPSO

It is straightforward to adapt the 2D method for estimating representativity error for data assimilation applications as most of the calculations can be performed offline. All that is required to compute during the assimilation

is the standard deviation of measurements within each gridbox, which are then multiplied by the appropriate scale factor. The correlation function of CloudSat radar reflectivity tends to predominantly be a function of latitude and height (Marchand, 2012). It therefore makes sense to compute the scale factor as a function of these indicators.

Figure 3.16 shows the scale factor as a function of longitude and latitude assuming a grid spacing of 18km, calculated using one year of CloudSat data. Interestingly the scale factor tends to be greatest in the tropics, apart from regions associated with stratocumulus decks off the west coast of Africa and the west coast of South America. The magnitude of the scale factor depends on two competing effects. Firstly, in regions of decreasing correlation, the variability of the measurements within a gridbox increases, thus increasing the representativity error. However, as variability within a gridbox increases, the number of effective samples along the measurement track increases, which decreases the representativity error. The dominant effect on controlling regional differences in scale factor changes with gridbox size.
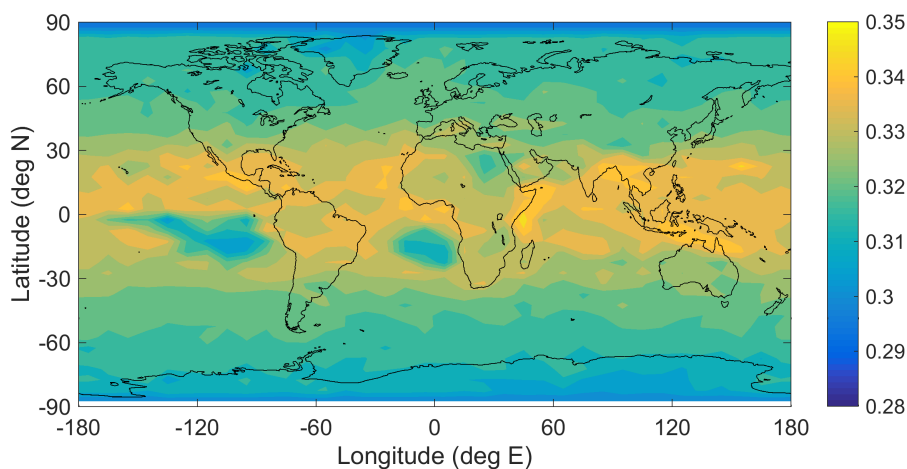

Figure 3.16: Spatial variability of the scale factor for the 2D method of estimating representativity error of radar reflectivity measurements. The correlation function was calculated using one year of CloudSat data. The scale factor is calculated at 1km intervals in height between 2-15 km and then averaged.

Figure 3.17 shows how the scaling factor varies as a function of grid spacing in two different regions. In the tropics, where the correlation between measurements tends to be less, for gridbox sizes between 10-50 km, the scaling factor is greater than the mid-latitudes. For larger grid boxes the opposite is true. The 'tipping point', where the two competing effects of gridbox size and track length are equal, appears to be for a grid box size of around 70 km.

3.3.4   Combining the observation errors

Now that we have defined the constituent parts of the observation errors, they can be combined to compute the total observation errors. Figure 3.18 shows the different sources of error along a transect covering different meteorological regimes. The observation operator error tends to be regionally uniform with a value around 3 dB. In areas of extreme precipitation the uncertainty increases, partly due to the compounding effect of uncertainties in attenuation. The representativity error shows a strong latitudinal variation, with areas of stratiform ice cloud in the mid-latitudes having the least error (around 2 dB). The error increases in tropical convection to over 6 dB. The measurement error is negligible compared to other errors in most cases apart from possibly near cloud edge, where it can reach around 2 dB. The combined observation error retains most of its structure from the representativity error, with the observation operator error tending to smooth the error field.
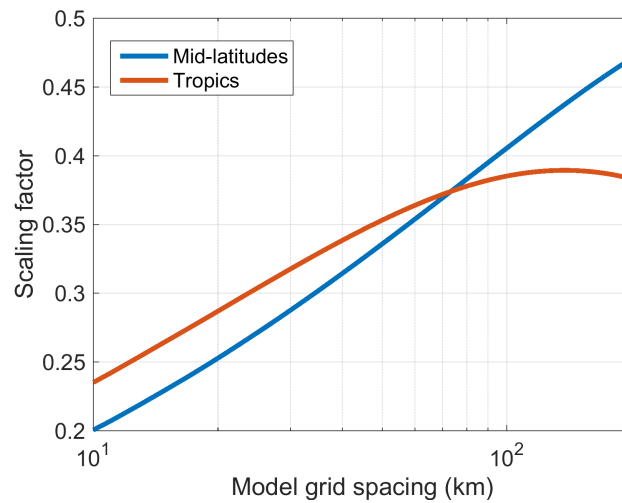
Figure 3.17: Scaling factor as a function of model grid spacing for correlations typical of radar reflectivity in the mid-latitudes (blue) and the tropics (red).
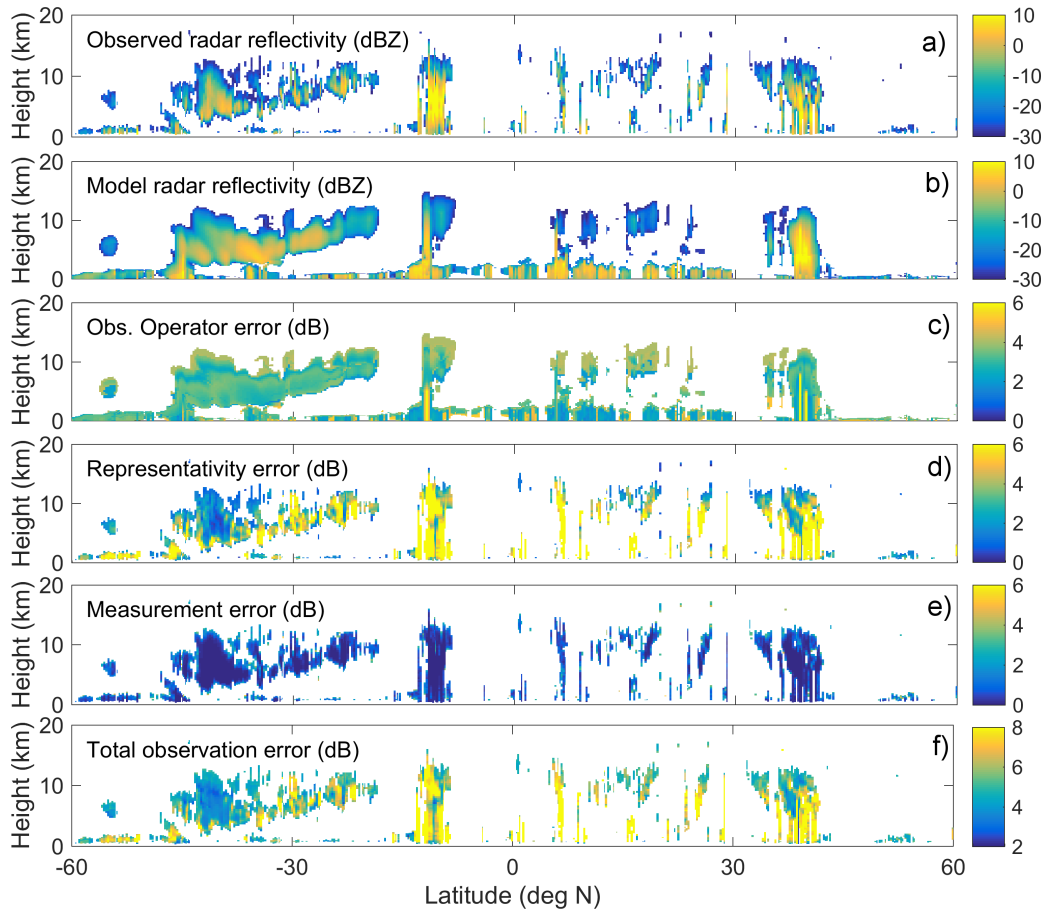

Figure 3.18: Example transect showing the different sources of observation error for CloudSat measurements between 0100-0130 UTC 1st August 2007. (a) shows the observed radar reflectivity, (b) forward modelled radar reflectivity, (c) observation operator error, (d) representativity error, (e) measurement error and (f) the total observation error.

An example of the combined observation error for CALIPSO is given in Fig. 3.19. Unlike for CloudSat measurements, the observation operator error tends to dominate for CALIPSO measurements, particularly where the signal begins to be attenuated. For liquid clouds, seen frequently along the transect at between 2-5 km, the operator error quickly exceeds 6 dB after the first few cloudy model layers. The representativity error has less latitudinal variation than for CloudSat, although areas with stratiform ice cloud in the mid-latitudes do tend to have lower error values. The measurement error for CALIPSO does become significant for areas of low backscatter, for example in the high-altitude ice-cloud above the equator.
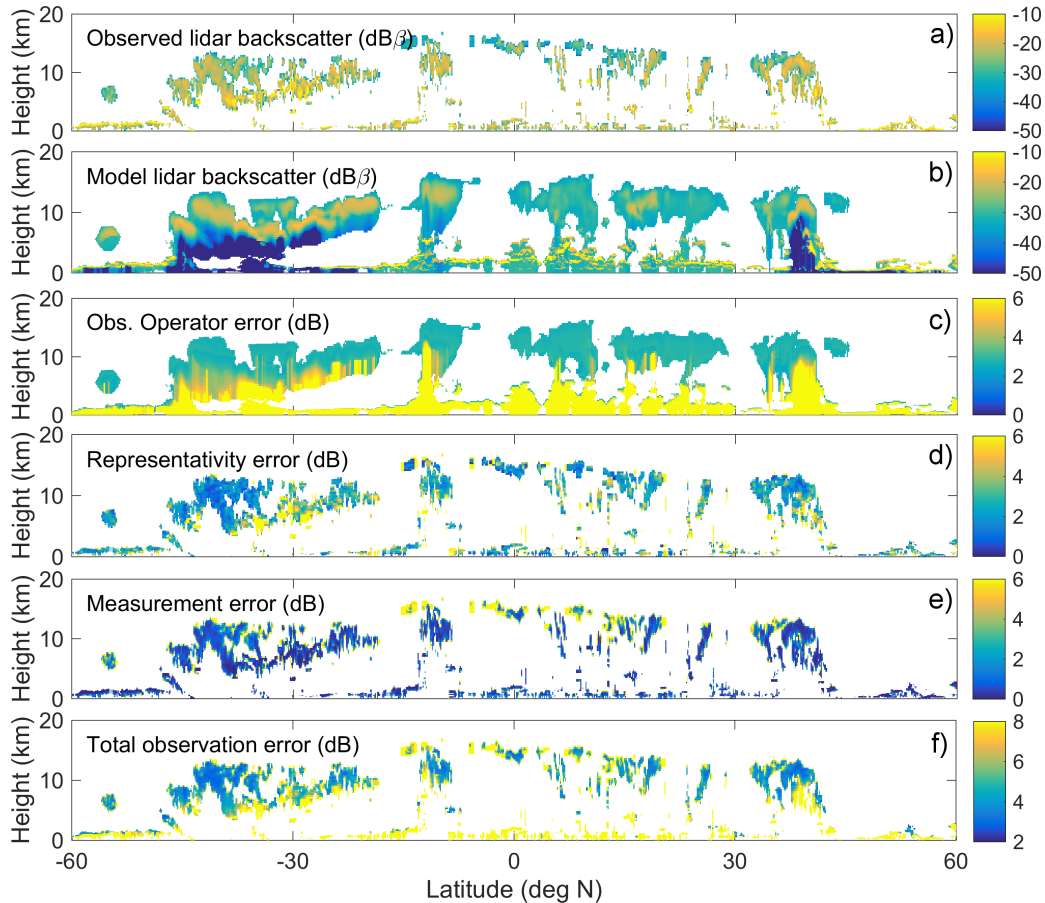

Figure 3.19: Same as Fig. 3.18, but for CALIPSO observations of attenuated lidar backscatter.

3.4   Global statistics

One way to test the validity of our observation error definitions is to compare them against the variability in first guess departures. It can be shown that the covariance of FG departures, **d**, is given by (e.g. Andersson, 2004):

$$< \mathbf{d}, \mathbf{d}^T > = \mathbf{HPH}^T + \mathbf{R} \tag{3.23}$$

where **H** is the tangent linear observation operator and **P** is the forecast error at observation time. If we assume $\mathbf{HPH}^T$ is of the order or smaller than **R**, then **R** should share some of the statistical characteristics of the first guess departures.

In a qualitative sense, the spatial pattern of observation errors for CloudSat and CALIPSO observations tend to agree with the corresponding standard deviations of first guess departures (Fig. 3.20). For CloudSat, the observation error shows a clear latitudinal dependence, which is also seen in the statistics of first guess departures.

The pattern of increased representativity error in the tropics (red ovals in Fig. 3.20) is also apparent in the FG departures. For CALIPSO, the converse is seen; areas of decreased observation error are seen in the tropics, perhaps a results of decreased variability in the ice cloud produced by convective outflow.
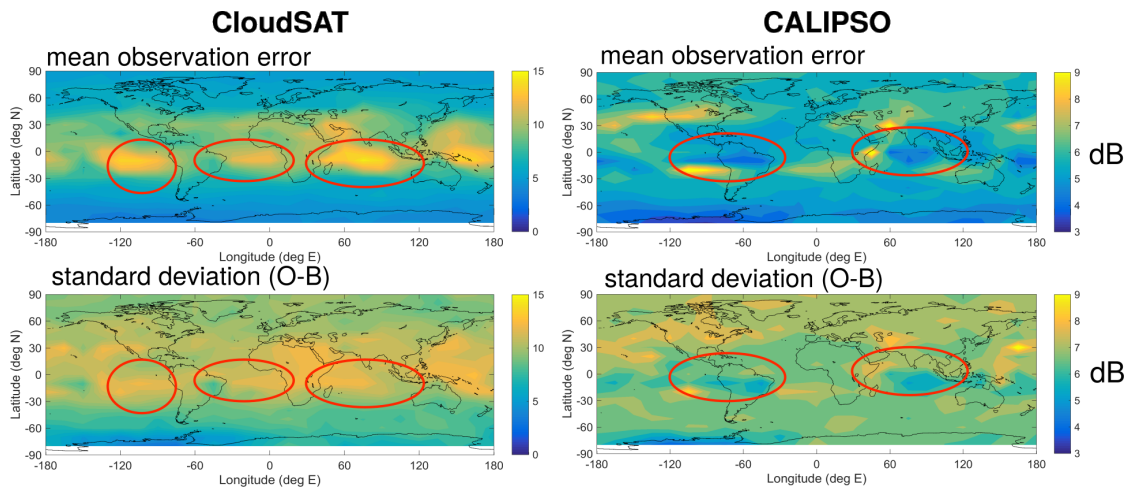

Figure 3.20: A comparison of global maps of CloudSat radar (left side) and CALIPSO lidar (right side) mean expected observation errors (sum of instrument, operator and representativity errors; top) versus the standard deviation of first guess departure errors (bottom). The red ovals are to aid comparison.

We can also compare the predicted observation error with the standard deviation of first guess departures in a more quantitative sense. By binning FG departures by the predicted observation error and taking the standard deviation, a clear correlation between FG departures and observation error can be seen for both CloudSat and CALIPSO observations (blue lines in Fig. 3.21). For small predicted observation error, the standard deviation of FG departures appears to converge to a value of around 6 dB. Given that the standard deviation of FG departures contains a contribution from forecast error, and if we assume the forecast error to be uniform across all weather regimes, we can estimate the forecast error to be around 6 dB for CloudSat radar reflectivity. Taking the same approach to CALIPSO observations leads to a forecast error of around 4 dB. If we subtract these estimates of forecast error from the standard deviation of first guess departures, the correlation between observation error and the residual FG departure error becomes even greater (red lines in Fig. 3.21).
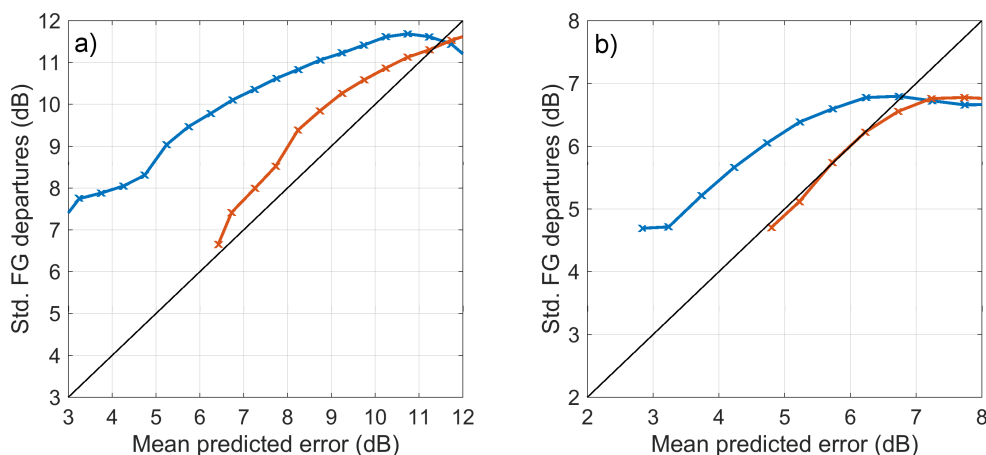

Figure 3.21: Standard deviation of first guess departures (blue) binned by predicted observation error during August 2007 for (a) CloudSat and (b) CALIPSO. Also shown (red) is the standard deviation of first guess departures minus an estimate of forecast error, see text for details.

Finally, we can also evaluate the distribution of FG departures, normalised by the predicted observation error. Looking at Fig. 3.22, we can see that normalising the FG departures by the observation error improves the gaussianity of FG departures (one of the assumptions of 4D-Var) for both CloudSat and CALIPSO observations. For CloudSat observations, normalising by the observation error improves the tails of the distribution out to three standard deviations. The tails of CALIPSO FG departure distribution are brought even closer to Gaussian when normalising by the observation error, suggesting that the assimilation will have an enhanced performance when using our predicted observation error, rather than a climatological average.
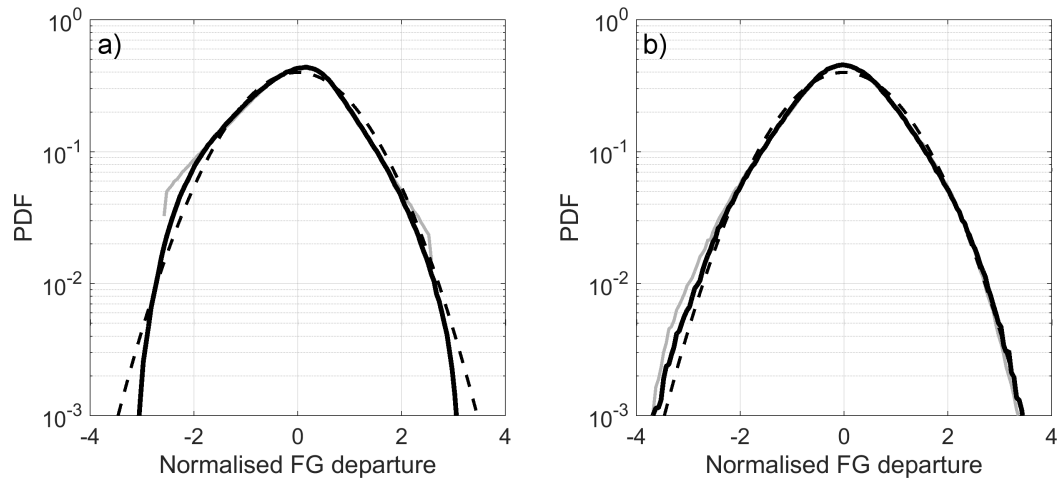

Figure 3.22: PDFs of (a) CloudSat and (b) CALIPSO FG departures normalised by the standard deviation of the whole sample (light grey line) or by the predicted observation error (black line). The dashed line shows the Gaussian distribution.

4   Data monitoring system

The monitoring of observational data against an NWP model's output is a fundamental step of quality control before performing data assimilation since it helps identify problems with observations and/or model. It also provides a template to understand and to exploit the new observations in an optimal way, before they become fully active in the analysis system. A data monitoring system also gives feedback to instrument mentors on potential measurement issues in a timely manor. As we will see, coupling observational information with model information allows for the quicker detection of errors than using observational information alone.

4.1   Description of the automatic monitoring system

All observations included in the ECMWF data assimilation system are included in an automatic data monitoring system. Four times per day (00, 06, 12, 18 UTC), corresponding to the operational early delivery assimilation system, selected statistical parameters are checked against expected ranges. Alerts are automatically triggered if statistics fall outside of their expected range. Two kinds of ranges are used by the automatic checking: 'soft' and 'hard' limits. Soft limits are updated automatically based on the statistics of the measurements in the past 20 days. Hard limits are set manually, and can be adjusted as needed. A severity level is also assigned to each alert depending on the magnitude of the anomaly.

Statistical 'indicators', whose ranges are monitored, are chosen to represent different aspects of an observation. In the ECMWF automatic monitoring system these indicators are typically, but not exclusively, one of the following:

- mean first guess departures
- standard deviation of first guess departures
- average analysis increment
- standard deviation of analysis increments
- number of observations

These indicators can also be subset by height or geolocation. To test the automatic system's capability to monitor EarthCARE data, we will use historical CloudSat and CALIPSO observations. Full functionality of the automatic system requires all the technical data handling developments and assimilation system developments (both reported in WP-3000, Janisková et al., 2017). However, in this work-package we propose and test indicators for observation and FG-related variables offline. We will also suggest definitions for the hard limits of the automatic monitoring system, and show how alerts are generated.

4.2   Radar

Before observations are included in the automatic monitoring system, we must first define hard limits for each indicator that will be used. One way to do this is by examining a climatology of the data (assumed to be free of any errors) and setting the limits using some threshold in the standard deviation. Figure 4.1 shows such a climatology using 3 months of CloudSat data averaged to the model resolution and corresponding model equivalents. The model equivalents are generated using the same model setup as in Sec. 3.2. The mean and standard deviation of each indicator are calculated with outliers removed in a two-step process; firstly all the data is included, then secondly any outliers are discarded and the statistics are re-calculated.

Analysing the histograms in Fig. 4.1, it is apparent that all the indicators follow quasi-Gaussian distributions. The closer the indicators' distributions are to the Gaussian distribution, the more suitable the data is for moni-

toring and detecting errors. The number of cloudy observations (Fig. 4.1c) contains some outliers; this could be due to drop-outs in measurements during the climatology period for re-pointing or satellite manoeuvres. One important point is that the standard deviation when using mean FG departures (0.288, see Fig. 4.1d) is significantly less than when using the mean of the observations alone (0.702, see Fig. 4.1a). This suggests that any drift in the observations would be detected first in the monitoring of first guess departures, thus stressing the importance of the model information in observation monitoring.
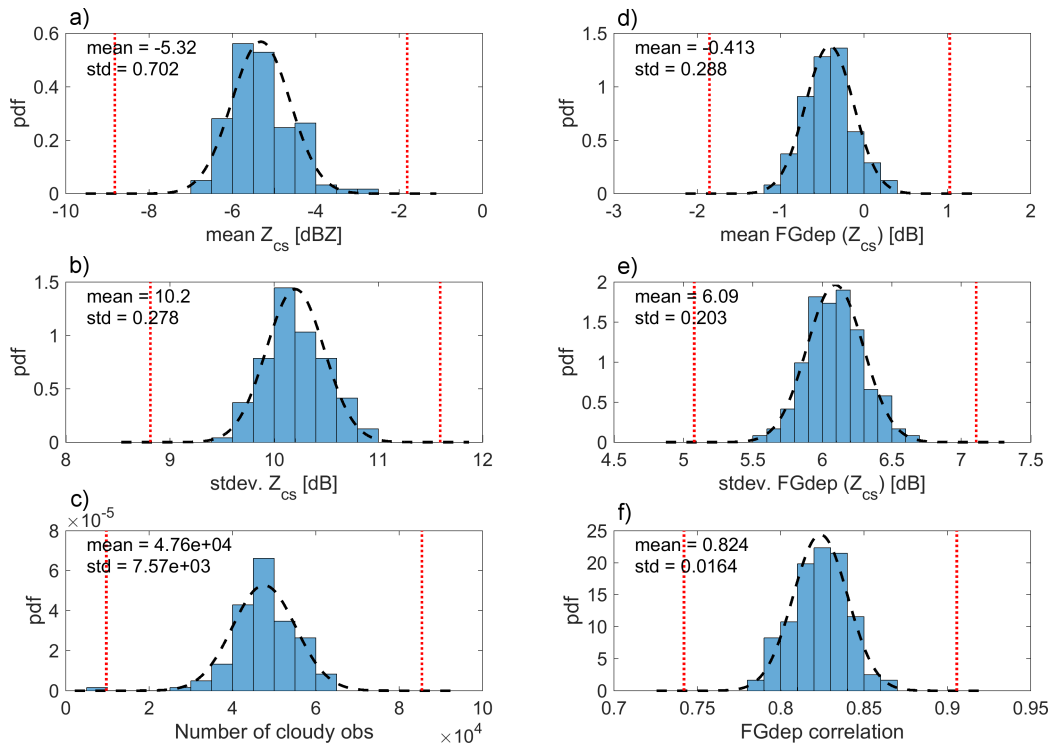

Figure 4.1: Histograms of 12-hour global mean CloudSat and model equivalent related variables for a three month period between August - September 2007. Panels on left side show observation related variables; (a) mean radar reflectivity, (b) standard deviation of radar reflectivity, (c) total number of cloudy observations, and panels on right side show observation and model related variables: (d) mean first guess departures, (e) standard deviation of first guess departures, (f) correlation between observed and model radar reflectivity. The black dashed line in each panel shows the Gaussian distribution with the mean and standard deviation of the data. The red dotted lines indicate 5 standard deviations from the mean.

Examples of the automated system using the indicators analysed in Fig. 4.1 are given in Figs. 4.2 and 4.3. All the indicators generally remain within the soft limits (green dashed lines), which are calculated using a running 20-day mean and a range of 3 standard deviations. One exception is for the number of cloudy observations and the global mean radar reflectivity around day 35, where a drop-out in the data has caused alerts to be triggered.

To understand the skill of the monitoring system to detect a problem in the quality of observations, experiments with artificially degraded CloudSat data are performed. Experiments are performed for the period of two months where the CloudSat radar calibration is assumed to drift after day 10. The drift is set to a 1% decrease per day, which leads to a total bias of 3 dB after two months. We compare the monitoring of stand-alone observations to the monitoring of observation and model related variables to investigate whether there are any advantages in considering FG departures compared to using CloudSat observations alone.

Figure 4.4 shows the monitoring of observation only indicators, where the bias has been introduced. Although a drift can be seen by eye in the global mean radar reflectivity (Fig. 4.4a), no additional alerts are triggered compared to the control (Fig. 4.2). However, when using the global mean FG departures (Fig. 4.5a), the drift
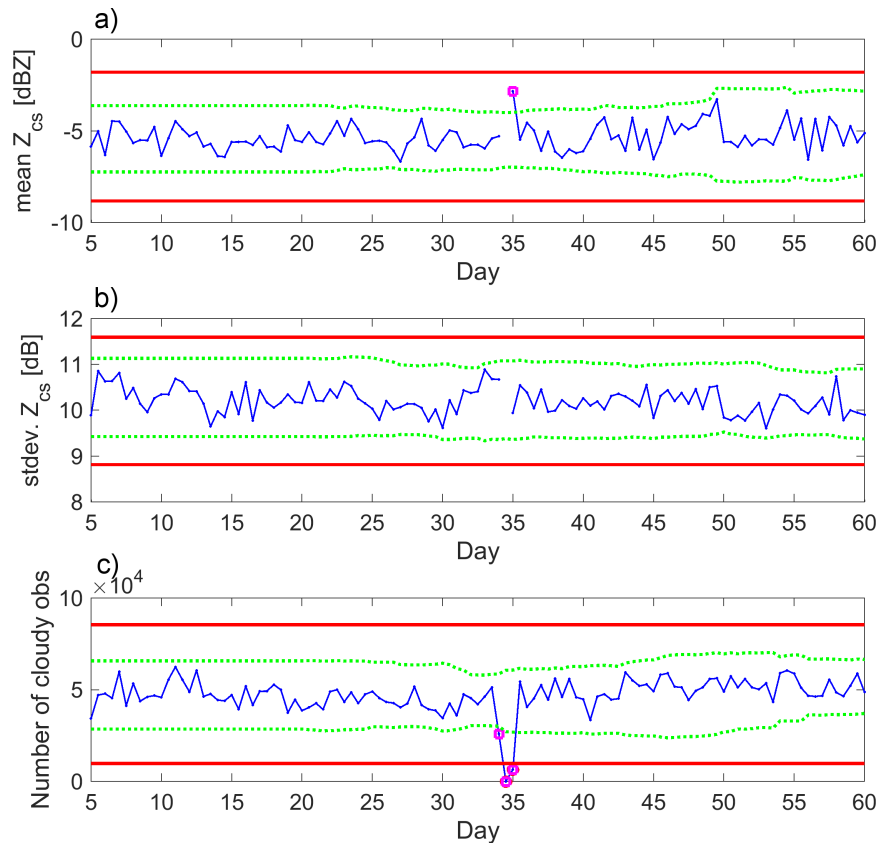
Figure 4.2: Example of CloudSat data within the automatic monitoring system using observation-only indicators of (a) global mean radar reflectivity, (b) global standard deviation of radar reflectivity and (c) total number of cloudy observations. The hard limits are given by the red lines and the soft limits are given by the green dashed lines. Alerts are shown by magenta circles.

is detected with alerts triggered around 30 days after the bias was introduced. Because the drift is gradual, the alerts are triggered by the hard limits rather than the soft limits. The soft limits are more likely to detect any sudden jumps in calibration or instrument issues. Although the drift is clearly visible in the mean, as might be expected, it does not affect the standard deviation of first guess departures or the correlation between observations and model.

4.3 Lidar

The same experiments and analysis is undertaken for CALIPSO observations of lidar attenuated backscatter. Similar to the Fig. 4.1, the global three-month statistics for different indicators are shown in Fig. 4.6. Again, the different variables follow a quasi-Gaussian distribution suggesting they are suitable for use in an automatic monitoring system. The standard deviation of global mean first guess departures (0.228, see 4.7a) is also similarly less than the mean of observations alone (0.292, see 4.7d). The reduction in correlation between observations and model equivalent (Fig. 4.7f) compared to CloudSat observations, explains why the relative difference between the standard deviation of mean first guess departures and the standard deviation of the mean of observations is not reduced by as much.

Examples of the automated system using the indicators analysed in Fig. 4.6 are given in Figs. 4.7 and 4.8. As seen with the radar observations, all the indicators generally remain within the soft limits (green dashed lines),
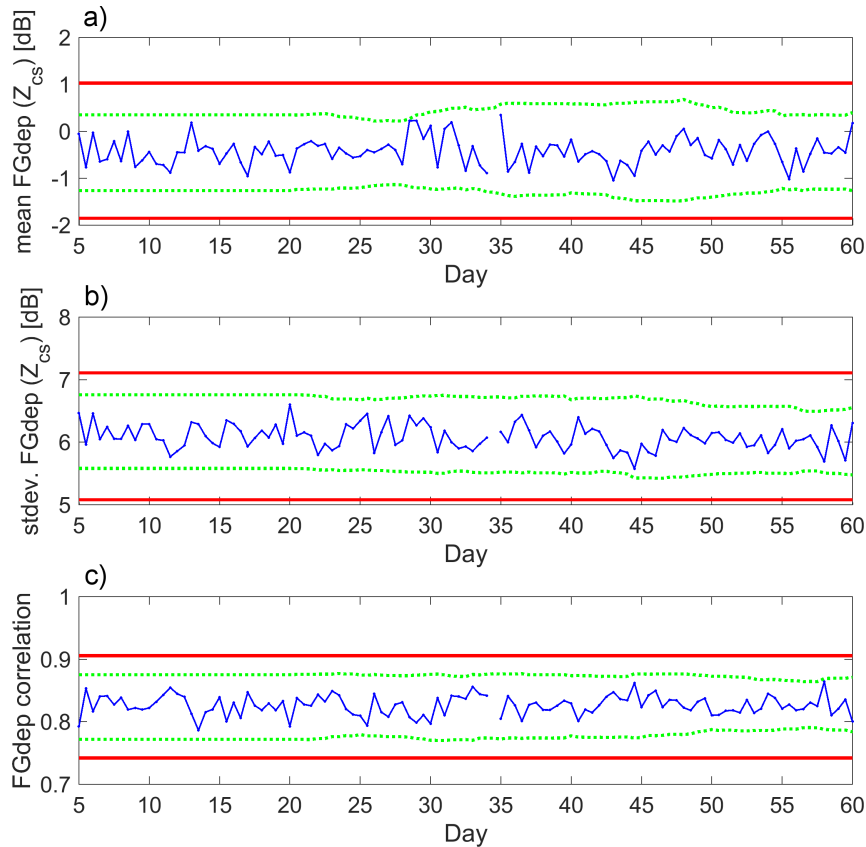
Figure 4.3: Example of CloudSat data within the automatic monitoring system using combined observation and model indicators of (a) global mean first guess departures, (b) global standard deviation of first guess departures and (c) correlation between observed and modelled radar reflectivity. The hard limits are given by the red lines and the soft limits are given by the green dashed lines. Alerts are shown by magenta circles.

which are calculated using a running 20-day mean and a range of 3 standard deviations. The soft-limit alert is triggered on two occasions when the mean of FG departures is used as an indicator (Fig. 4.8b), but these isolated alarms are likely just to have occurred by chance, rather than for any specific issue or malfunction of the instrument. There could also be some issue with the model at these occasions; in the full in-line monitoring system this could be checked by looking at other observations being monitored.

As shown for the monitoring of CloudSat observations, figures 4.9 and 4.10 show an example of the automatic monitoring system where an artificial drift has been introduced to the observations. Again, the experiments are performed for the period of two months, but this time the CALIPSO lidar calibration is assumed to drift after day 10. The drift is also set to a 1% decrease per day, which leads to a total bias of 3 dB after two months.

Due to the narrower dynamic range of CALIPSO attenuated backscatter compared to CloudSat radar reflectivity, the artificial drift in calibration is first detected after only 20 days in observations alone (Fig. 4.9a), although only seven alarms are triggered for the total 60 day period. In contrast, using the mean FG departure as an indicator (Fig. 4.10a) results in an alarm triggered in less than 15 days after the bias is introduced. By the end of the 60 day period a total of over thirty alerts would have been issued. As seen when monitoring radar reflectivity, the drift does not affect the other indicators.
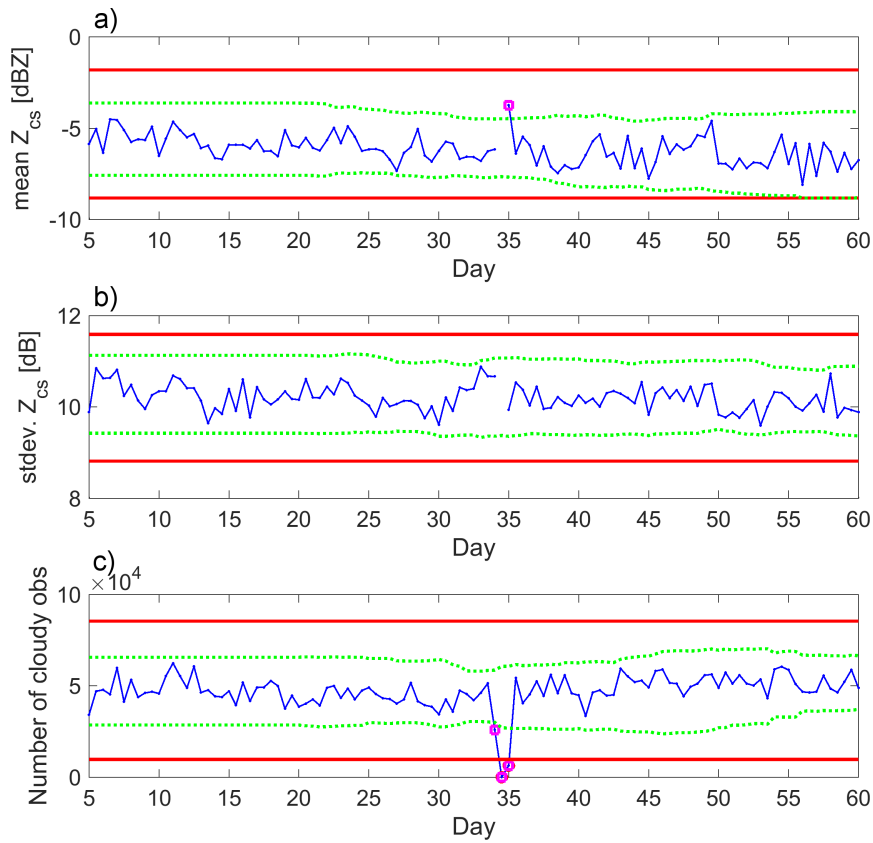
Figure 4.4: Same as Fig. 4.2, but where a 1% per day drift in observed radar reflectivity has been introduced at day 10.
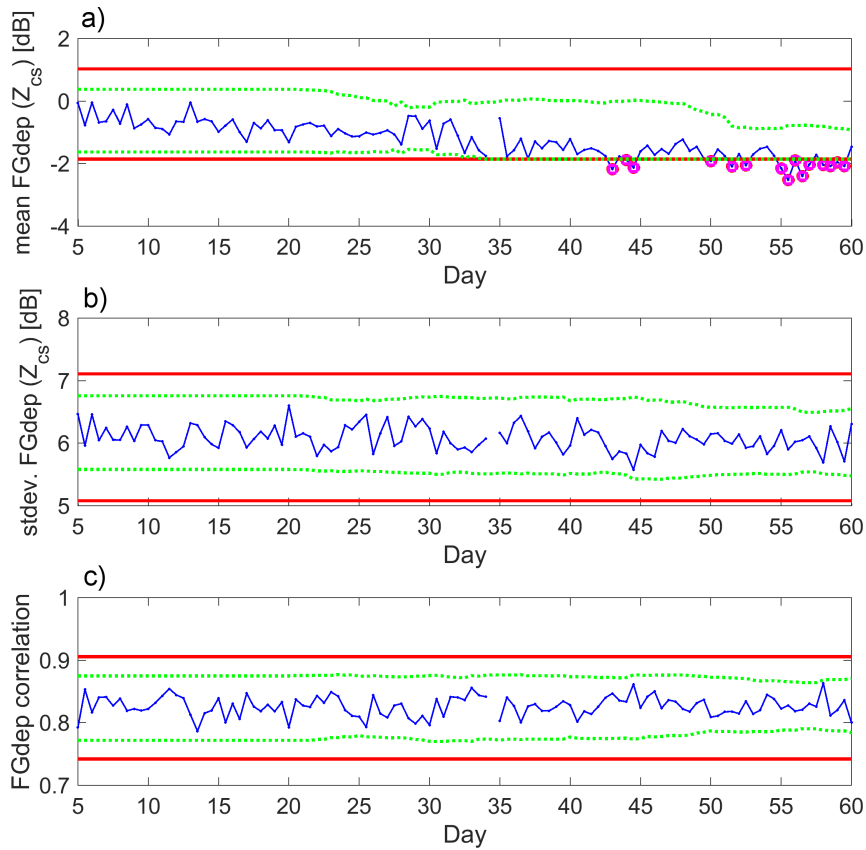

Figure 4.5: Same as Fig. 4.3, but where a 1% per day drift in observed radar reflectivity has been introduced at day 10.
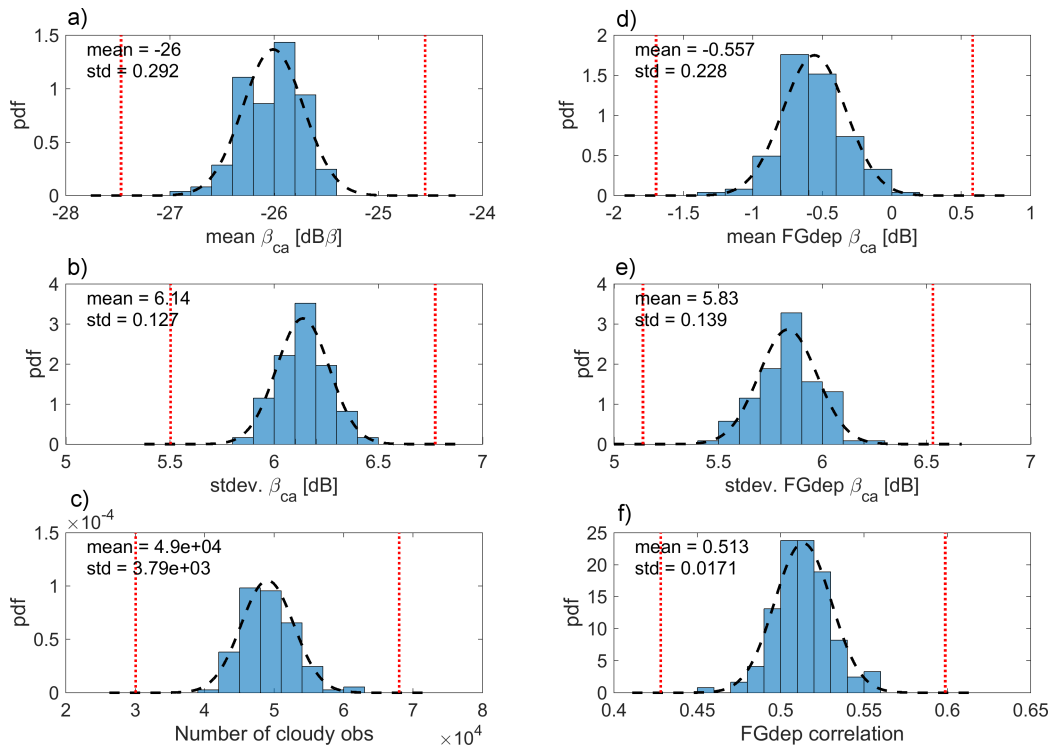
Figure 4.6: Histograms of 12-hour global mean CALIPSO and model equivalent related variables for a three month period between August - September 2007. Panels on left side show observation related variables; (a) mean lidar attenuated backscatter, (b) standard deviation of attenuated backscatter, (c) total number of cloudy observations, and panels on right side show observation and model related variables: (d) mean first guess departures, (e) standard deviation of first guess departures, (f) correlation between observed and model radar reflectivity. The black dashed line in each panel shows the Gaussian distribution with the mean and standard deviation of the data. The red dotted lines indicate 5 standard deviations from the mean.
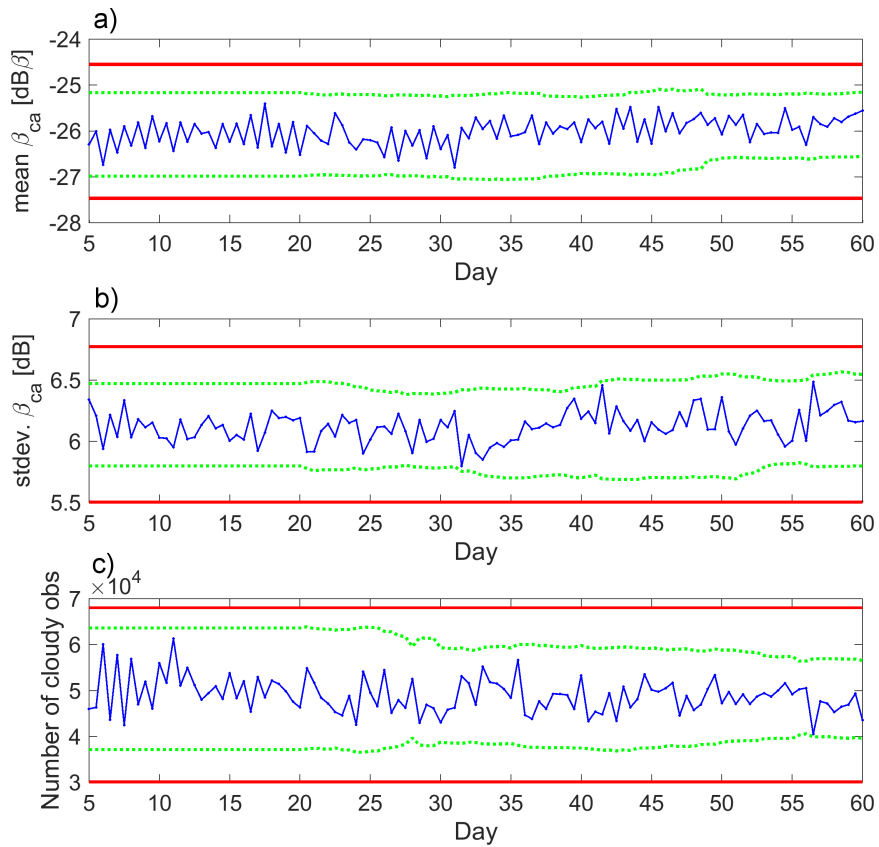
Figure 4.7: Example of CALIPSO data within the automatic monitoring system using observation-only indicators of (a) global mean lidar attenuated backscatter, (b) global standard deviation of lidar backscatter and (c) total number of cloudy observations. The hard limits are given by the red lines and the soft limits are given by the green dashed lines. Alerts are shown by magenta circles.

Figure 4.8: Example of CALIPSO data within the automatic monitoring system using combined observation and model indicators of (a) global mean first guess departures, (b) global standard deviation of first guess departures and (c) correlation between observed and modelled radar reflectivity. The hard limits are given by the red lines and the soft limits are given by the green dashed lines. Alerts are shown by magenta circles.
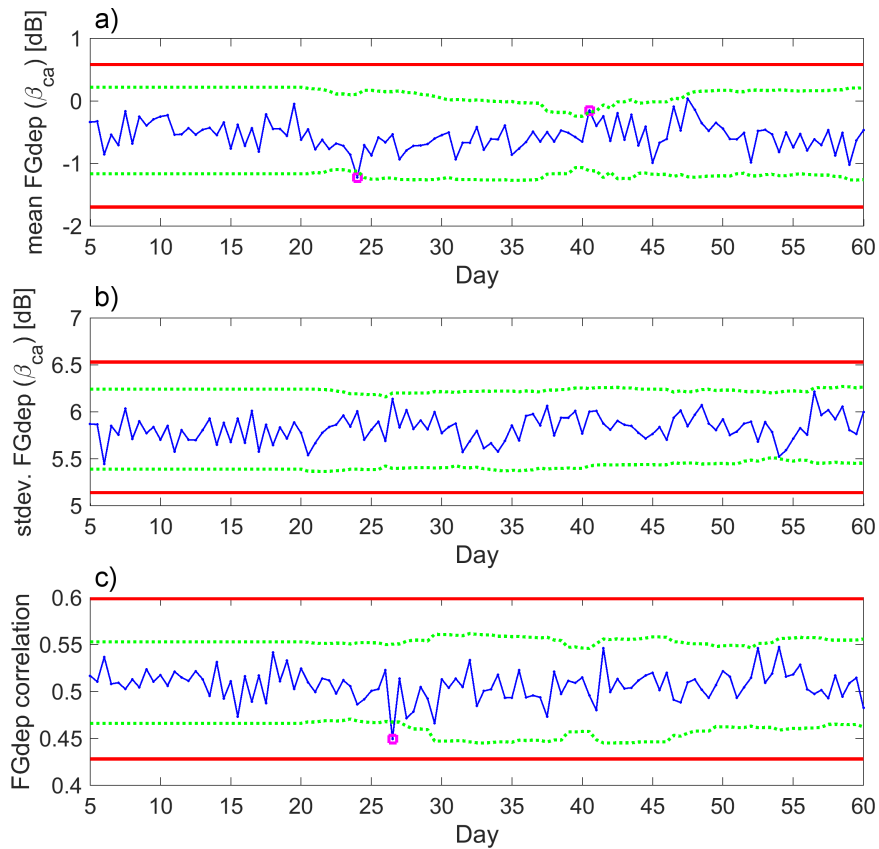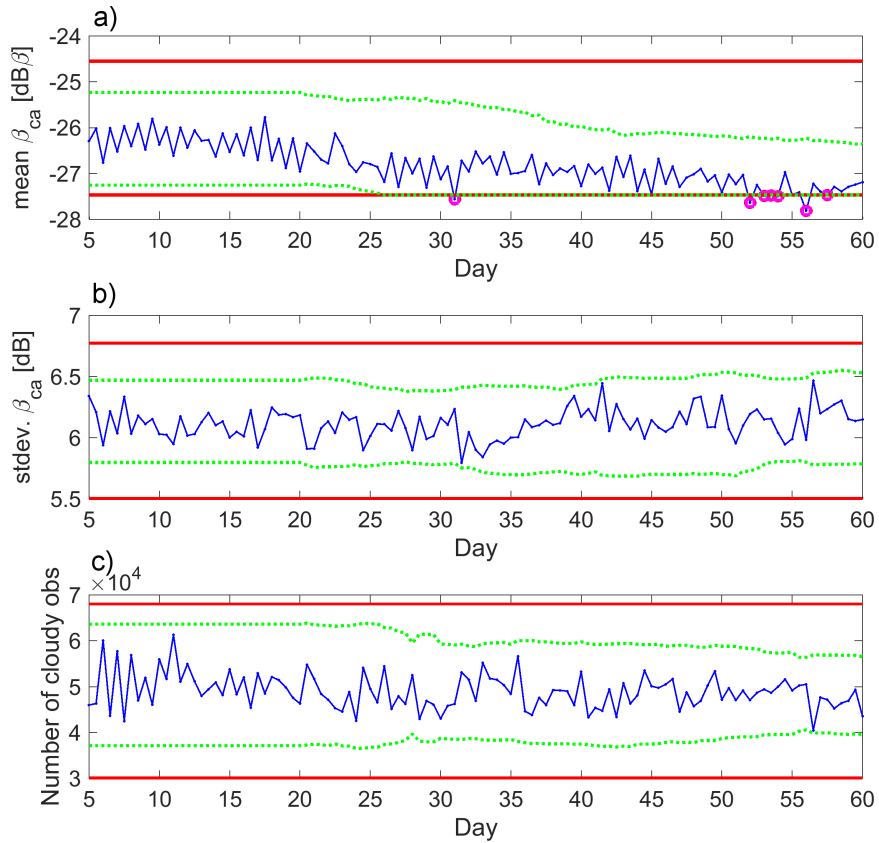
Figure 4.9: Same as Fig. 4.7, but where a 1% per day drift in observed lidar attenuated backscatter has been introduced at day 10.
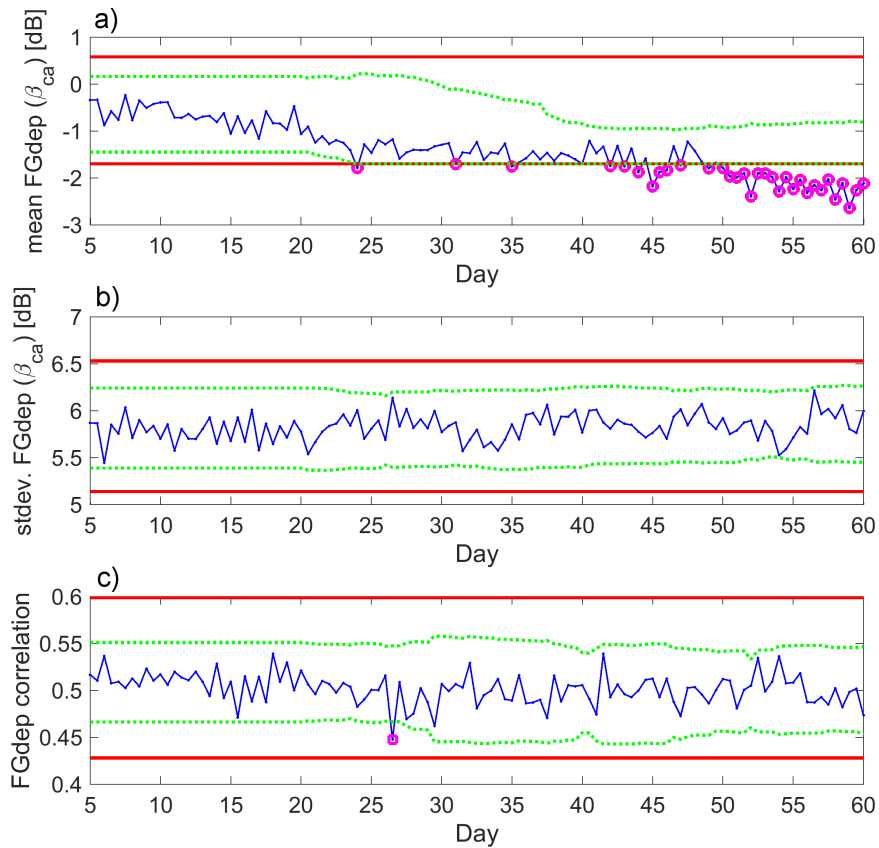

Figure 4.10: Same as Fig. 4.8, but where a 1% per day drift in observed lidar attenuated backscatter has been introduced at day 10.

5   Summary and conclusions

In this report we have detailed the necessary scientific developments in preparation for the direct assimilation of cloud radar and lidar satellite observations into the ECMWF 4D-Var system. As explained in the introduction, data assimilation fundamentally relies on three pieces of information: the observations, the model equivalent of the observations and the expected error of the two. While the technical developments related to the handling of the actual observations will be given in WP-3000 (Assimilation system development for cloud radar and lidar observations, Janisková et al., 2017), this workpackage focuses on updating the observation operator and the observation errors with the latest research. We also introduce the automatic monitoring system of the observations.

Specifically, the microphysical assumptions in the observation operator were updated to provide physical consistency with the IFS cloud scheme. An alternative approach to account for cloud overlap was also implemented and demonstrated to provide similar results to the multi-column approach at a fraction of the computational cost. Additional computational cost savings were achieved by introducing a parameterized version of the scattering look up tables and general streamlining of the code.

The expected observation error was characterised by taking an 'error inventory' approach; we define the measurement error, operator error and representativity error explicitly before combining them to give the overall observation error. In particular, a new flow-dependent representativity error method was introduced and shown to account for the observation/model mis-match in spatial scales almost perfectly in an idealised situation and with a similar performance to the SFM method (Stiller, 2010) when using 2D MODIS data of cloud optical depth. To specify the uncertainty in the observation operator due to unrepresented microphysical variability, a Monte-Carlo approach was taken where hydrometeor PSD parameters were perturbed within physical ranges.

Another important task for successful data assimilation is screening and bias correction of observations. Following on from previous work during QuARL (Janisková et al., 2010) and STSE (Janisková et al., 2014), a bias correction scheme using the updated model configuration and customised indicators has been implemented. The bias correction scheme was then tested using independent data. As with the observation operator and error characterisation, some updates to the bias correction scheme may be necessary using knowledge gained from the feasibility studies that will be performed in WP-5000 (Feasibility demonstration of 4D-Var assimilation system using CloudSat and CALIPSO observations, Janisková and Fielding, 2018). A review of the screening of the observations was made and the percentage of observations discarded for various thresholds was shown.

Finally, the automatic monitoring system was introduced and demonstrated off-line. Statistics of various indicators were generated for both radar and lidar using CloudSat and CALIPSO data so that hard limits in the automatic monitoring system can be set. Tests of the system using an artificial drift in the observations showed that the detection of errors was greatly improved by monitoring observations and model information simultaneously compared to observations alone. This highlights the advantage of monitoring observations at an operational NWP centre over monitoring observations at their source.

List of Acronyms

1D-Var	One-Dimensional Variational Assimilation
4D-Var	Four-Dimensional Variational Assimilation
AD	ADjoint
ATLID	ATmospheric LIDar
C-PRO	Cloud profiling radar PROcessing
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CloudSat	NASA's cloud radar mission
CF	Cloud Fraction
CPR	Cloud Profiling Radar
CPU	Central Processing Unit
EarthCARE	Earth, Clouds, Aerosols and Radiation Explorer
ECMWF	European Centre for Medium Range Weather Forecasts
EDA	Ensemble of Data Assimilations
ESA	European Space Agency
ESRAD	Edward and Slingo RADiation code
ESTEC	European Space Research and Technology Centre
FG	First Guess
FOV	Field Of View
IFS	Integrated Forecasting System of ECMWF
JAXA	Japan Aerospace eXploration Agency
MODIS	MOderate-resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NL	Non-Linear
NSF	Noise Scale Factor
NWP	Numerical Weather Prediction
ODB	Observation Data Base
PSD	Particle Size Distribution
PVC	Photon Variance-Covariance
QuARL	Quantitative Assessment of the Operational Value of Space-Borne Radar and Lidar Measurements of Cloud
rmse	root-mean square error
SCOPS	Sub-grid Cloud Overlap Profile Sampler
SFM	Structure Function Maximum
STSE	Support-to-Science-Element
TCo639	Model cubic octahedral grid with spectral truncation T639
TDTS	Time-Dependent Two-Stream
TL	Tangent Linear
TRMM	Tropical Rainfall Measurement Mission
UTC	Universal Time Coordinated
VarBC	Variational Bias Correction
Z	Radar reflectivity
ZmVar	Z (reflectivity) Model for Variational assimilation at ECMWF
WP	Work Package

References

Abel, S. J. and I. A. Boutle, 2012: An improved representation of the raindrop size distribution for single-moment microphysics schemes, *Quarterly Journal of the Royal Meteorological Society*, **138**(669), 2151–2162.

Anderson, T. W., 1994: *The Sample Mean, Covariances, and Spectral Density*, John Wiley & Sons, Inc., pp. 438–500, ISBN 9781118186428.

Andersson, E., 2004: Modelling the temporal evolution of innovation statistics, in *ECMWF seminar proceedings: recent developments in data asimilation for atmosphere and ocean, 8-12 Sept., 2003*, Eur. Cent. for Med. Range Weather Forecasts, Reading, UK, available from http://www.ecmwf.int, pp. 153–164.

Aulign, T., A. P. McNally, and D. P. Dee, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system, *Quarterly Journal of the Royal Meteorological Society*, **133**(624), 631–642.

Battaglia, A., T. Augustynek, S. Tanelli, and P. Kollias, 2011: Multiple scattering identification in space-borne w-band radar measurements of deep convective cores, *Journal of Geophysical Research: Atmospheres*, **116**(D19), n/a–n/a, D19201.

Bauer, P., A. J. Geer, P. Lopez, and D. Salmond, 2010: Direct 4D-Var assimilation of all-sky radiances: Part I. Implementation, *Quart. J. Roy. Meteorol. Soc.*, **136**, 1868–1885.

Benedetti, A., P. Lopez, P. Bauer, and E. Moreau, 2006: Experimental use of TRMM precipitation radar observations in 1D+4D-Var assimilation, *Q. J. R. Meteorol. Soc.*, **131**, 2473–2495.

Collis, R. and P. Russell, 1976: Lidar measurement of particles and gases by elastic backscattering and differential absorption, *Laser monitoring of the atmosphere*, pp. 71–151.

Dee, D., 2004: Variational bias correction of radiance data in the ECMWF system, in *ECMWF workshop proceedings: Assimilation of high spectral resolution sounders in NWP, 28 June – 1 July, 2004*, Eur. Cent. for Med. Range Weather Forecasts, Reading, UK, available from http://www.ecmwf.int, pp. 97–112.

Delanoë, J. and R. J. Hogan, 2010: Combined cloudsat-calipso-modis retrievals of the properties of ice clouds, *Journal of Geophysical Research: Atmospheres*, **115**(D4), n/a–n/a, D00H29.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space, *Quart. J. Roy. Meteorol. Soc.*, **131**, 3385–3396.

Di Michele, S., M. Ahlgrimm, R. Forbes, M. Kulie, R. Bennartz, M. Janisková, and P. Bauer, 2012: Iterpreting and evaluation of the ECMWF global model with CloudSat onbservations: ambiguities due to radar reflectivity forwad operator uncertainties, *Q. J. R. Meteorol. Soc.*, **138**, 2047–2065, doi:10.1002/qj.1936.

Di Michele, S., E. Martins, and M. Janisková, 2014a: Observation operator and observation processing for cloud radar, WP-1100 report for the project Support-to-Science-Element STSE Study - EarthCARE Assimilation, AO/1-6441/10/NL/CB, 59 pp.

Di Michele, S., E. Martins, and M. Janisková, 2014b: Observation operator and observation processing for cloud lidar, WP-1200 report for the project Support-to-Science-Element STSE Study - EarthCARE Assimilation, AO/1-6441/10/NL/CB, 40 pp.

Edwards, J. M. and A. Slingo, 1996: Studies with a flexible new radiation code. i: Choosing a configuration for a large-scale model, *Quarterly Journal of the Royal Meteorological Society*, **122**(531), 689–719.

Field, P., A. Heymsfield, and A. Bansemer, 2007: Snow size distribution parametrization for midlatituide and tropical ice cloud,, *J. Atmos. Sci.*, **64(12)**, 4346–4365.

Fielding, M., M. Janisková, and R. Hogan, 2017: EarthCARE data handling and testing, WP-4000 report for the project Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, ESA ESTEC contract 4000116891/16/NL/LvH, 29 pp.

Geer, A. J. and P. Bauer, 2011: Observation errors in all-sky data assimilation, *Quart. J. Roy. Meteorol. Soc.*, **137**, 2024–2037.

Geer, A. J., P. Bauer, and P. Lopez, 2010: Direct 4D-Var assimilation of all-sky radiances: Part II. Assessment, *Quart. J. Roy. Meteorol. Soc.*, **136**, 1886–1905.

Geleyn, J. F. and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation, *Contrib. Atmos. Phys.*, **52**, 1–16.

Heymsfield, A. J., S. Lewis, A. Bansemer, J. Iaquinta, L. M. Miloshevich, M. Kajikawa, C. Twohy, and M. R. Poellot, 2002: A general approach for deriving the properties of cirrus and stratiform ice cloud particles, *Journal of the Atmospheric Sciences*, **59**(1), 3–29.

Heymsfield, G. M., 1976: Statistical objective analysis of dual-doppler radar data from a tornadic storm, *Journal of Applied Meteorology*, **15**(1), 59–68.

Hildebrand, P. H. and R. S. Sekhon, 1974: Objective determination of the noise level in doppler spectra, *Journal of Applied Meteorology*, **13**(7), 808–811.

Hogan, R. J., 2008: Fast lidar and radar multiple-scattering models. part i: Small-angle scattering using the photon variancecovariance method, *Journal of the Atmospheric Sciences*, **65**(12), 3621–3635.

Hogan, R. J. and A. Battaglia, 2008: Fast lidar and radar multiple-scattering models. part ii: Wide-angle scattering using the time-dependent two-stream approximation, *Journal of the Atmospheric Sciences*, **65**(12), 3636–3651.

Hogan, R. J. and A. J. Illingworth, 2000: Deriving cloud overlap statistics from radar, *Quarterly Journal of the Royal Meteorological Society*, **126**(569), 2903–2909.

Hogan, R. J., M. P. Mittermaier, and A. J. Illingworth, 2006: The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model, *Journal of Applied Meteorology and Climatology*, **45**(2), 301–317.

Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus A*, **38**, 111–136.

Illingworth, A. et al., 2015: The earthcare satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation, *Bulletin of the American Meteorological Society*, **96**(8), 1311–1332.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF, *ECMWF tech. memo.*, **636**, available from www.ecmwf.int.

Janisková, M., S. Di Michele, and E. Martins, 2014: Support-to-Science-Elements (STSE) Study - EarthCARE Assimilation, ESA Contract Report on Project 4000102816/11/NL/CT, 225 pp.

Janisková, M. and M. Fielding, 2018: Feasibility demonstration of 4D-Var assimilation system using CloudSat and CALIPSO observations, WP-5000 report for the project Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, ESA ESTEC contract 4000116891/16/NL/LvH, 35 pp.

Janisková, M., M. Fielding, M. Crepulja, D. Vasiljević, T. Král, and P. Lean, 2017: Assimilation system development for cloud radar and lidar observations, WP-3000 report for the project Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction, ESA ESTEC contract 4000116891/16/NL/LvH, 26 pp.

Janisková, M. and P. Lopez, 2013: Linearized physics for data assimilation at ECMWF, in *S.K. Park and L. Xu (Eds), Data Assimilation for Atmospheric, Ocean and Hydrological Applications (Vol II), Springer-Verlag Berline Heidelberg*, pp. 251–286, doi:10.1007/978–3–642–35088–7–11.

Janisková, M., P. Lopez, and P. Bauer, 2012: Experimental 1D+4D-Var assimilation of CloudSat observations, *Q. J. R. Meteorol. Soc.*, **138**, 1196–1220, doi:10.1002/qj.988.

Janisková, M., O. Stiller, S. Di Michele, R. Forbes, J.-J. Morcrette, M. Ahlgrimm, P. Bauer, and L. Jones, 2010: QuARL - Quantitative Assessment of the Operational Value of Space-Borne Radar and Lidar Measurements of Cloud and Aerosol Profiles, ESA Contract Report on Project 21613/08/NL/CB, 329 pp.

Kitchen, M., 1989: Representativeness errors for radiosonde observations, *Quarterly Journal of the Royal Meteorological Society*, **115**(487), 673–700.

Kollias, P., A. Battaglia, A. Tatarevic, W. Szrymer, and D. Burns, 2016: Doppler radar and synergy products for EarthCARE Algorithm Theoretical Basis Document, Unpublished.

Kollias, P., E. E. Clothiaux, M. A. Miller, B. A. Albrecht, G. L. Stephens, and T. P. Ackerman, 2007: Millimeter-wavelength radars: New frontier in atmospheric cloud and precipitation research, *Bulletin of the American Meteorological Society*, **88**(10), 1608–1624.

Kulie, M. S., R. Bennartz, T. J. Greenwald, Y. Chen, and F. Weng, 2010: Uncertainties in microwave properties of frozen precipitation: implications for remote sensing and data assimilation, *J. Atmos. Sci.*, **67**, 3471–3487.

Liang, X.-Z. and W.-C. Wang, 1997: Cloud overlap effects on general circulation model climate simulations, *Journal of Geophysical Research: Atmospheres*, **102**(D10), 11039–11047.

Liebe, H. J., 1985: An updated model for millimeter wave propagation in moist air, *Radio Science*, **20**(5), 1069–1089.

Liebe, H. J., P. W. Rosenkranz, and G. A. Hufford, 1992: Atmospheric 60-ghz oxygen spectrum - new laboratory measurements and line parameters, *jqsrt*, **48**, 629–643.

Liu, G., 2008: A database of microwave single-scattering properties for nonspherical ice particles, *Bulletin of the American Meteorological Society*, **89**(10), 1563–1570.

Liu, Z., W. Hunt, C. Vaughan, C. Hostetler, M. McGill, K. Powell, D. Winker, and Y. Hu, 2006: Estimating random errors due to shot noise in backscatter lidar observations, *Applied Optics*, **45(18)**, 4437–4447.

Locatelli, J. D. and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles, *Journal of Geophysical Research*, **79**(15), 2185–2197.

Marchand, R., 2012: Spatial correlation of hydrometeor occurrence, reflectivity, and rain rate from cloudsat, *Journal of Geophysical Research: Atmospheres*, **117**(D6), n/a–n/a, D06202.

Miles, N. L., J. Verlinde, and E. E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds, *Journal of the Atmospheric Sciences*, **57**(2), 295–311.

Mitchell, D. L., 1996: Use of mass- and area-dimensional power laws for determining precipitation particle terminal velocities, *Journal of the Atmospheric Sciences*, **53**(12), 1710–1723.

Morcrette, J.-J. and Y. Fouquart, 1986: The overlapping of cloud layers in shortwave radiation parametrizations, *J. Atmos. Sci.*, **43**, 321–328.

Pardo-Igúzquiza, E. and M. Chica-Olmo, 1993: "the fourier integral method: An efficient spectral method for simulation of random fields", *"Mathematical Geology"*, **25**(2), 177–217.

Pardo-Igúzquiza, E. and M. Chica-Olmo, 1994: Specsim: A program for simulating random fields by an improved spectral approach, *Computers and Geosciences*, **20**, 597–613.

Platt, C., 1973: Lidar and radioinetric observations of cirrus clouds, *Journal of the Atmospheric Sciences*, **30**(6), 1191–1204.

Shonk, J. P. and R. J. Hogan, 2008: Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1d radiation schemes by using three regions at each height, *Journal of Climate*, **21**(11), 2352–2370.

Smythe, G. R. and D. S. Zrnic, 1983: Correlation analysis of doppler radar data and retrieval of the horizontal wind, *Journal of Climate and Applied Meteorology*, **22**(2), 297–311.

Stiller, O., 2010: A flow-dependent estimate for the sampling error, *J. Geophys. Res.*, **115 (D22)**, doi: 10.1029/2010JD013934.

Webb, M., C. Senior, S. Bony, and J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Climate Dynamics*, **17**(12), 905–922.

Yang, P., K. N. Liou, K. Wyser, and D. Mitchell, 2000: Parameterization of the scattering and absorption properties of individual ice crystals, *Journal of Geophysical Research: Atmospheres*, **105**(D4), 4699–4718.